

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

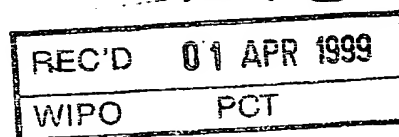
Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTC.

EJAU
**PRIORITY
DOCUMENT**SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)**Bescheinigung**

Die BASF Aktiengesellschaft in Ludwigshafen/Deutschland hat
eine Patentanmeldung unter der Bezeichnung


"Verfahren zur Herstellung von Biotin"

am 19. Februar 1998 beim Deutschen Patentamt eingereicht.

Die angehefteten Stücke sind eine richtige und genaue
Wiedergabe der ursprünglichen Unterlagen dieser Patent-
anmeldung.

Die Anmeldung hat im Deutschen Patentamt vorläufig die Sym-
bole C 12 N, C 07 H und C 12 P der Internationalen Patent-
klassifikation erhalten.

München, den 29. September 1998
Der Präsident des Deutschen Patentamts
Im Auftrag



Aktenzeichen: 198 06 872.7

Rixner

Patentansprüche

1. Verfahren zur Herstellung von Biotin, dadurch gekennzeichnet,
5 daß man ein S-Adenosyl-Methionin-Synthasegen mit der Sequenz
SEQ ID No. 1 und mindestens ein weiteres Biotin-Biosynthese-
gen bioS1, bioS2 oder bioS3 mit den Sequenzen SEQ ID No. 3,
SEQ ID No. 5 oder SEQ ID No. 7 sowie ihre funktionellen Va-
10 rianten, Analoge oder Derivate in einem prokaryontischen oder
eukaryontischen Wirtsorganismus, der in der Lage ist Biotin
zu synthetisieren, exprimiert, diesen züchtet und das synthe-
tisierte Biotin direkt, nach Abtrennung der Biomasse oder
nach Reinigung des Biotins verwendet.
- 15 2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß es
sich bei den Varianten der Gene mit den Sequenzen SEQ ID
No.1, SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 um Gene
handelt, die auf der von den Sequenzen nach Anspruch 1 abge-
20 leiteten Aminosäureebene eine Homologie von 30 bis 100 % auf-
weisen und eine gesteigerte Biotinsynthese ermöglichen.
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß
als Wirtsorganismus ein Organismus ausgewählt aus der Gruppe
25 der Gattungen Escherichia, Citrobacter, Serratia, Klebsiella,
Salmonella, Pseudomonas, Comamonas, Acinetobacter, Azotobac-
ter, Chromobacterium, Bacillus, Clostridium, Arthrobacter,
Corynebacterium, Brevibacterium, Lactococcus, Lactobacillus,
Streptomyces, Rhizobium, Agrobacterium, Staphylococcus, Rho-
30 dotorula, Sporobolomyces, Yarrowia, Schizosaccharomyces oder
Saccharomyces verwendet wird.
4. Verfahren nach den Ansprüchen 1 bis 3, dadurch gekennzeich-
35 net, daß als Wirtsorganismus eine regulationsdefekte Biotin-
mutante verwendet wird.
5. Verfahren nach den Ansprüchen 1 bis 4, dadurch gekennzeich-
net, daß die Expression mindestens einer Kopie der Gene mit
den Sequenzen SEQ ID No.1, SEQ ID No. 3, SEQ ID No. 5 und SEQ
ID No. 7 nach Anspruch 1 allein oder mit einer oder mehreren
40 Kopien mindestens eines weiteren Biotingens ausgewählt aus
der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW,
bioX, bioY oder bioR in einem prokaryontischen oder eukaryon-
tischen Wirtsorganismus erfolgt.

UP

2

6. Verfahren nach den Ansprüchen 1 bis 5, dadurch gekennzeichnet, daß die Expression mindestens einer Kopie der Gene mit den Sequenzen SEQ ID No. 1, SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 nach Anspruch 1 allein oder mit einer oder mehreren Kopien mindestens eines weiteren Biotingens ausgewählt aus der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR in einem prokaryontischen oder eukaryontischen Wirtsorganismus auf einem gemeinsamen oder auf getrennten Vektoren erfolgt.
7. Genkonstrukt enthaltend ein S-Adenosyl-Methionin-Synthasegen mit der SEQ ID No. 1 und mindestens ein weiteres Biotin-Biosynthesegen bioS1, bioS2 oder bioS3 mit den Sequenzen SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 sowie ihre funktionellen Varianten, Analoge oder Derivate, das mit einem oder mehreren Regulationssignalen zur Erhöhung der Genexpression und/oder Proteinexpression funktionell verknüpft ist und/oder dessen natürliche Regulation ausgeschaltet wurde.
8. Genkonstrukt nach Anspruch 7, dadurch gekennzeichnet, daß es in einem Vektor inseriert wurde, der für die Expression des Gens in einem prokaryontischen oder eukaryontischen Wirtsorganismus geeignet ist.
9. Genkonstrukt nach Anspruch 7 oder 8, dadurch gekennzeichnet, daß die Gene mit den Sequenzen SEQ ID No. 1, SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 sowie ihre funktionellen Varianten, Analoge oder Derivate in mehreren Kopien im Genkonstrukt vorliegen.
10. Genkonstrukt nach Ansprüchen 7 bis 9, dadurch gekennzeichnet, daß das S-Adenosyl-Methionin-Synthasegen SEQ ID No. 1 und mindestens ein weiteres Biotin-Biosynthesegen bioS1, bioS2 oder bioS3 mit den Sequenzen SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 sowie ihre funktionellen Varianten, Analoge oder Derivate nach Anspruch 7 zusammen mit einer oder mehreren Kopien mindestens eines weiteren Gens ausgewählt aus der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR im Genkonstrukt oder Vektor vorliegt.
- ~~11. Organismen enthaltend ein Genkonstrukt gemäß den Ansprüchen 7 bis 10.~~
12. Verwendung der Sequenzen gemäß Anspruch 1 zur Herstellung von Biotin.

3

13. Verwendung des bioS3-Gens mit der Sequenz SEQ ID No. 7 seiner funktionellen Varianten, Analoge oder Derivate allein oder in Kombination mit mindestens einem weiteren Gen ausgewählt aus der Gruppe S-Adenosyl-Methionin-Synthasegen, bioS1, bioS2, bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR zur Herstellung von Biotin.

14. Verwendung eines Genkonstrukts gemäß den Ansprüchen 7 bis 10 zur Herstellung von Biotin.

10

15

20

25

30

35

40

45

Verfahren zur Herstellung von Biotin

Beschreibung

5

Die Erfindung betrifft ein Genkonstrukt, enthaltend ein S-Adenosyl-Methionin-Synthasegen, mit der Sequenz SEQ ID No. 1 und ein Biotin-Biosynthesegen bioS1, bioS2 und/oder bioS3 mit den Sequenzen SEQ ID No.3, SEQ ID No.5 bzw. SEQ ID No.7 und gegebenenfalls
10 mindestens einer weiteren Biotinsynthesegensequenz ausgewählt aus der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR. Die Erfindung betrifft weiterhin Organismen, die dieses Genkonstrukt enthalten und die Verwendung des Genkonstrukts zur Herstellung von Biotin, sowie ein Verfahren zur Herstellung von Biotin.
15

Als Coenzym spielt Biotin (Vitamin H) eine essentielle Rolle bei enzymkatalysierten Carboxylierungs- und Decarboxylierungsreaktionen. Biotin ist damit ein essentieller Faktor in lebenden Zellen.
20 Biotin muß von fast allen Tieren und einigen Mikroorganismen von außen aufgenommen werden, da sie Biotin nicht selber synthetisieren können. Es ist damit für diese Organismen ein essentielles Vitamin. Bakterien, Hefen und Pflanzen hingegen können Biotin aus Vorstufen selbst synthetisieren (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326, DeMoll, E., Escherichia coli and Salmonella, eds. Neidhardt, F. C. et al. ASM Press, Washington DC, USA, 1996: 704 - 708, ISBN 1-55581-084-5).
25

Die Biotinsynthese wurde in bakteriellen Organismen speziell im gramnegativen Bakterium Escherichia coli und im grampositiven Bakterium Bacillus sphaericus untersucht (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326). Als erstes bisher bekanntes Intermediat in E. coli wird Pimelyl-CoA (PmCoA) angesehen, das aus der Fettsäuresynthese stammt (DeMoll, E., Escherichia coli and Salmonella, eds. Neidhardt, F. C. et al. ASM Press, Washington DC, USA, 1996: 704 - 708, ISBN 1-55581-084-5 1996).
30 Der Syntheseweg dieser Biotin-Vorstufe in E. coli ist bisher weitgehend unbekannt (Lemoine et al., Mol. Microbiol. 19, 1996: 645 - 647). Es wurden mit bioC und bioH zwei Gene identifiziert, deren korrespondierende Proteine für die Synthese von Pm-CoA verantwortlich sind. Die enzymatische Funktion der Genprodukte BioH und BioC ist bisher nicht bekannt (Lemoine et al., Mol. Microbiol. 19, 1996: 645 - 647, DeMoll, E., Escherichia coli and Salmonella, eds. Neidhardt, F. C. et al. ASM Press, Washington DC, USA, 1996: 704 - 708, ISBN 1-55581-084-5). Pm-CoA wird in vier
35 weiteren enzymatischen Schritten zum Biotin umgewandelt. Ausgehend vom Pm-CoA findet zuerst die Kondensation mit Alanin zu
40
45

7-Keto-8-Amino-Pelargonsäure (KAPA) durch BioF statt. Es folgt die Umsetzung von KAPA zu 7,8 Diamino-Pelargonsäure (DAPA) durch BioA. Der nächste Schritt führt nach einer ATP-abhängigen Carboxylierungsreaktion zum Dethiobiotin (DTB) und wird durch BioD katalysiert. Im letzten Schritt findet die Umsetzung von DTB zu Biotin statt. Dieser Schritt wird durch BioB katalysiert. Der chemische und enzymatische Mechanismus der Umwandlung von DTB zu Biotin ist bisher nur unvollständig verstanden und aufgeklärt.

- 10 Eine Charakterisierung der Umwandlung von DTB zu Biotin wurde bisher ausschließlich in bakteriellen bzw. pflanzlichen Zellextrakten durchgeführt (WO94/8023, EP-B-0 449 724, Sanyal et al. Arch. Biochem. Biophys., Vol. 326, No. 1, 1996: 48 - 56 und Biochemistry 33, 1994: 3625 - 3631, Baldet et al. Europ. J. Biochem. 217, 1, 1993: 479 - 485, Méjean et al. Biochem. Biophys. Res. Commun., Vol. 217, No. 3, 1995: 1231 - 1237, Ohshiro et al., Biosci. Biotechnol. Biochem., 58, 9, 1994: 1738 - 1741).

- 20 In vitro Studien haben gezeigt, daß niedermolekulare Faktoren wie NADPH, Cystein, Thiamin, Fe^{2+} , Asparagin, Serin, Fruktose 1-6-bisphosphat und S-Adenosylmethionin die Synthese von Biotin stimulieren können (Ohshiro et al., Biosci. Biotechnol. Biochem., 58, 9, 1994: 1738 - 1741, Birch et al., J. Biol. Chem. 270, 32, 1995: 19158 - 19165, Ifuku et al., Biosci. Biotechnol. Biochem., 59, 2, 1995: 185 - 189, Sanyal et al. Arch. Biochem. Biophys. 326, 1, 1996: 48 - 56).

- 30 Neben diesen niedermolekularen Faktoren wurden weitere Proteine identifiziert, die die Synthese von Biotin aus DTB in Gegenwart von BioB stimulieren. Dabei handelt es sich um Flavodoxin und Flavodoxin-NADPH-Reduktase (Birch et al., J. Biol. Chem. 270, 32, 1995: 19158 - 19165, Ifuk et al., Biosci. Biotechnol. Biochem., 59, 2, 1995: 185 - 189, Sanyal et al., Arch. Biochem. Biophys. 326, 1, 1996: 48 - 56). Weitere Proteine, die eine Stimulation der Biotinsynthese bewirken, sind die in der deutschen Anmeldung mit der Anmeldenummer 197.31274.8 (Priorität 22.7.97) beschriebenen Gene bioS1 und bioS2.

- 40 Über die Herkunft des Schwefels im Biotinmolekül existieren unterschiedliche Ergebnisse. Bei Untersuchungen der Biotinsynthese in Gesamtzellextrakten wurde gezeigt, daß in Gegenwart von ^{35}S -markiertem Cystein Radioaktivität in Biotin inkorporiert wurde; weder mit ^{35}S -markiertem Methionin noch mit S-Adenosyl-Methionin konnte ein Schwefel-Einbau ins Biotinmolekül nachgewiesen werden (Ifuku et al., Biosci. Biotechnol. Biochem. 59, 2, 1995: 184 - 189, Birch et al., J. Biol. Chem. 270, 32, 1995: 19158 - 19165).

Die für die beschriebenen Proteine kodierenden Gene bioF, bioA,

3

- bioD, und bioB sind in *E. coli* auf einem bidirektionalem Operon kodiert. Dieses Operon liegt zwischen der λ -attachement-site und dem *uvrB* Gen Locus bei ca. 17 Minuten auf dem *E. coli* Chromosom (Berlyn et al. 1996: 1715 - 1902). Auf diesem Operon sind zusätzlich noch zwei weitere Gene kodiert, von denen das eine, bioC, bereits beschriebene Funktionen in der Synthese von Pm-CoA hat, während einem offenen Leseraster hinter bioA bisher keine Funktion zugeordnet werden konnte (WO94/8023, Otsuka et al., J. Biol. Chem. 263, 1988: 19577 - 85). Hoch konservierte Homologe zu den *E. coli* Proteinen BioF, A, D, B wurden in *B. sphaericus*, *B. subtilis*, *Syneccocystis* sp. (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326, Bower et al., J. Bacteriol. 175, 1996: 4122 - 4130, Kaneko et al., DNA Res. 3, 3, 1996: 109 - 136), Archaeabakterien wie *Methanococcus janaschi*, Hefen wie *Saccharomyces cerevisiae* (Zhang et al., Arch. Biochem. Biophys. 309, 1, 1994: 29 - 35) oder in Pflanzen wie *Arabidopsis thaliana* (Baldet et al., C. R. Acad. Sci. III, Sci. Vie. 319, 2, 1996: 99 - 106)) gefunden.
- Die Synthese von Pm-CoA scheint in den beiden bisher untersuchten gram-positiven Mikroorganismen anders zu verlaufen als in *E. coli*. Es konnten keine Homologen von bioH und bioC gefunden werden (Brown et al. Biotechnol. Genet. Eng. Rev. 9, 1991: 295 - 326).
- Biotin ist eine optisch aktive Substanz mit drei Chiralitätszentren. Es wird bisher wirtschaftlich nur über eine vielstufige, teure chemische Synthese hergestellt werden.
- Alternativ zu dieser chemischen Synthese wurde eine Vielzahl von Versuchen unternommen, ein fermentatives Verfahren zur Herstellung von Biotin mit Mikroorganismen aufzubauen. Durch Klonierung des Biotin-Operons auf multi-copy-Plasmide konnte die Biotinsynthese in den mit diesen Genen transformierten Mikroorganismen erhöht werden. Eine weitere Erhöhung der Biotinsynthese wurde durch die Deregulierung der Biotingenexpression über die Selektion von *birA*-Mutanten erreicht (Pai C. H., J. Bacteriol. 112, 1972: 1280 - 1287). Die Kombination beider Ansätze, das heißt die Expression der Plasmid-kodierten Biosynthesegene in einem regulationsdefizienten Stamm (EP-B-0 236 429), brachte eine weitere Steigerung der Produktivität. Das Biotin-Operon kann dabei entweder unter Kontrolle seines nativen bidirektionalem Promotors verbleiben (EP-B-0 236 429), oder seine Gene können unter die Kontrolle eines extern regulierbaren Promotors gebracht werden (WO94/08023).
- Durch die bisher verfolgten Ansätze zur fermentativen Herstellung von Biotin in *E. coli* konnte keine wirtschaftlich ausreichende

Produktivität erreicht werden.

Aufgabe der vorliegenden Erfindung war es, ein technisches, fermentatives Verfahren zur Herstellung von Biotin zu entwickeln,
5 daß eine möglichst hohe Biotinsynthese zeigt.

Diese Aufgabe wurde durch das erfindungsgemäße Verfahren zur Herstellung von Biotin, dadurch gekennzeichnet, daß man ein S-Adenosyl-Methionin-Synthase (SAM-Synthase)-Gen mit der Sequenz SEQ ID
10 No. 1 und mindestens ein weiteres Biotin Biosynthesegen bioS1, bioS2 oder bioS3 mit den Sequenzen SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 sowie ihre funktionellen Varianten, Analoge oder Derivate in einem prokaryontischen oder eukaryontischen Wirtsorganismus, der in der Lage ist Biotin zu synthetisieren, exprimiert,
15 diesen züchtet und das synthetisierte Biotin direkt, nach Abtrennung der Biomasse oder nach Reinigung des Biotins verwendet, gelöst.

Die im erfindungsgemäßen Verfahren verwendeten Gene, das SAM-Synthase-Gen mit der Sequenz SEQ ID No. 1 und die Biotin-Biosynthesegene bioS1, bioS2 und bioS3 mit den Sequenzen SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 werden in der SwissProt-Datenbank unter den "Accession"-Nummern P04384 (metK), U29581 (bioS1), P39171 (bioS2) und D90811 (bioS3) geführt. In der Datenbank sind
25 eine Reihe von Homologen zu MetK aus E. coli beschrieben. Diese Homologe umfassen Organismen wie weitere Eubakterien (z.B. H. influenza, B. subtilis), wie auch Eukaryonten (z.B. Hefen: S. cerevisiae, Planta: P. deltoides, Arthropoda D. melanogaster, und Mammalia: R. norvegicus).

30 Durch Expression einer oder mehrerer des SAM-Synthase-Gens mit der Sequenz SEQ ID No. 1, seiner funktionellen Varianten, Analoge oder Derivate in Kombination mit mindestens einem der Biotinsynthesegene bioS1, bioS2 oder bioS3 mit den Sequenzen SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 sowie deren funktionelle Varianten,
35 Analoge oder Derivate in einem prokaryontischen oder eukaryontischen Wirtsorganismus läßt sich die Produktivität der Biotinbiosynthese deutlich steigern. Bevorzugt wird eine Kombination von SAM-Synthase-Gen und bios1 zur Expression verwendet. Vorteilhafterweise wird für eine gesteigerte Biotinsynthese gleichzeitig
40 mindestens ein weiteres Biotingen ausgewählt aus der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR mit exprimiert. Durch die Expression der Gene wird die Synthese von Biotin um mindestens den Faktor 2 gegenüber der Kontrolle ohne diese Gene, bevorzugt um einen Faktor größer 3, gesteigert.
45

5

Nach Isolierung und Sequenzierung sind die im erfindungsgemäßen Verfahren verwendeten Gene, das SAM-Synthase-Gen mit der Nukleotidsequenz SEQ ID No. 1, das bioS1 Gen mit der Nukleotidsequenz SEQ ID No. 3, das bioS2 Gen mit der Nukleotidsequenz SEQ ID No. 5 und das bioS3 Gen mit der Nukleotidsequenz SEQ ID No. 7 erhältlich, die für die in SEQ ID NO: 2, respektive SEQ ID No. 4, respektive SEQ ID No. 6 und SEQ ID No. 8 angegebenen Aminosäuresequenzen oder deren Allelvariationen kodieren. Unter Varianten sind SEQ ID No. 1-, SEQ ID No. 3- bzw. SEQ ID No. 5, respektive SEQ ID No. 7-Varianten zu verstehen, die 30 bis 100 % Homologie auf Aminosäureebene, bevorzugt 50 bis 100 %, ganz besonders bevorzugt 80 bis 100 % aufweisen. Allelvarianten umfassen insbesondere funktionelle Varianten, die durch Deletion, Insertion oder Substitution von Nukleotiden aus den in SEQ ID NO: 1, SEQ ID No. 3, SEQ ID No. 5 und SEQ ID No. 7 dargestellten Sequenzen erhältlich sind, wobei die enzymatische Aktivität aber erhalten bleibt.

Weiterhin sind unter Varianten auch funktionelle Äquivalente der Gene wie die O-Acetyl-serinsulfohydrolase A, die O-Acetyl-serinsulfohydrolase B, die β -Cystathionase (siehe Flint et al., J. Biol. Chem., Vol. 271, 1996: 16053 - 16067) oder nifS und seine prokaryontischen und eukaryontischen Homologen beispielsweise aus Klebsiella, Candida, Hefen oder aus Caenorhabditis zu verstehen, die in der Lage sind die enzymatische Aktivität von bioS1, bioS2 oder bioS3 in der Biotinsynthese zu übernehmen.

Unter funktionellen Analogon von SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 und SEQ ID No. 7 sind beispielsweise ihre prokaryontischen oder eukaryontischen Homologen wie bakterielle, pilzliche, pflanzliche, tierische oder menschliche Homologen zu verstehen. Unter Analogon sind weiterhin auch verkürzte Sequenzen, Einzelstrang-DNA oder RNA der codierenden und nichtcodierenden DNA-Sequenz zu verstehen.

Derivate sind beispielsweise Promotorvarianten zu verstehen. Die Promotoren, die den angegebenen Nukleotidsequenzen vorgeschaltet sind, können durch ein oder mehrere Nukleotidaustausche, durch Insertion(en) und/oder Deletion(en) verändert sein, ohne daß aber die Funktionalität bzw. Wirksamkeit der Promotoren beeinträchtigt wird. Des weiteren können die Promotoren durch Veränderung ihrer Sequenz in ihrer Wirksamkeit erhöht oder komplett durch wirkksamere Promotoren auch artfremder Organismen ausgetauscht werden.

Unter Derivaten sind auch Varianten zu verstehen, deren Nukleotidsequenz im Bereich von -1 bis -30 vor dem Startkodon so verändert wurden, daß die Genexpression und/oder die Proteinexpression

erhöht wird. Vorteilhafterweise geschieht dies durch eine veränderte Shine-Dalgarno-Sequenz.

Als prokaryontische Wirtsorganismen des erfindungsgemäßen Verfahrens kommen prinzipiell alle Biotin synthetisierenden gram-negativen oder gram-positiven Bakterien in Frage. Als gram-negative Bakterien seien beispielhaft Enterobacteriaceae wie die Gattungen *Escherichia*, *Aerobacter*, *Enterobacter*, *Citrobacter*, *Shigella*, *Klebsiella*, *Serratia*, *Erwinia* oder *Salmonella*, *Pseudomonadaceae* wie die Gattungen *Pseudomonas*, *Xanthomonas*, *Burkholderia*, *Glucobacter*, *Nitrosomonas*, *Nitrobacter*, *Methanomonas*, *Comamonas*, *Cellulomonas* oder *Acetobacter*, *Azotobacteraceae* wie die Gattungen *Azotobacter*, *Azomonas*, *Beijerinckia* oder *Derxia*, *Neisseriaceae* wie die Gattungen *Moraxella*, *Acinetobacter*, *Kingella*, *Neisseria* oder *Branhamella*, die *Rhizobiaceae* wie die Gattungen *Rhizobium* oder *Agrobacterium* oder die gram-negativen Gattungen *Zymomonas*, *Chromobacterium* oder *Flavobacterium*, genannt. Als gram-positive Bakterien seien beispielhaft die Endosporen-bildenden gram-positiven aeroben oder anaeroben Bakterien wie die Gattungen *Bacillus*, *Sporolactobacillus* oder *Clostridium*, die coryneformen Bakterien wie die Gattungen *Arthrobacter*, *Cellulomonas*, *Curtobacterium*, *Corynebacterium*, *Brevibacterium*, *Microbacterium* oder *Kurtzia*, die *Actinomycetales* wie die Gattungen *Mycobacterium*, *Rhodococcus*, *Streptomyces* oder *Nocardia*, die *Lactobacillaceae* wie die Gattungen *Lactobacillus* oder *Lactococcus*, die gram-positiven Kokken wie die Gattungen *Micrococcus* oder *Staphylococcus*, genannt.

Bevorzugt werden Bakterien der Gattungen *Escherichia*, *Citrobacter*, *Serratia*, *Klebsiella*, *Salmonella*, *Pseudomonas*, *Comamonas*, *Acinetobacter*, *Azotobacter*, *Chromobacterium*, *Bacillus*, *Clostridium*, *Arthrobacter*, *Corynebacterium*, *Brevibacterium*, *Lactococcus*, *Lactobacillus*, *Streptomyces*, *Rhizobium*, *Agrobacterium* oder *Staphylococcus* im erfindungsgemäßen Verfahren verwendet. Besonders bevorzugt werden Gattungen und Arten wie *Escherichia coli*, *Citrobacter freundii*, *Serratia marcescens*, *Salmonella typhimurium*, *Pseudomonas mendocina*, *Pseudomonas aeruginosa*, *Pseudomonas mutabilis*, *Pseudomonas chlororaphis*, *Pseudomonas fluorescens*, *Comamonas acidovorans*, *Comamonas testosteroni*, *Acinetobacter calcoaceticus*, *Azotobacter vinelandii*, *Chromobacterium violaceum*, *Bacillus subtilis*, *Bacillus sphaericus*, *Bacillus stearothermophilus*, *Bacillus pumilus*, *Bacillus licheniformis*, *Bacillus amyloliquefaciens*, *Bacillus megaterium*, *Bacillus cereus*, *Bacillus thuringiensis*, *Arthrobacter citreus*, *Arthrobacter paraffineus*, *Corynebacterium glutamicum*, *Corynebacterium primorioxydans*, *Corynebacterium sp.*, *Brevibacterium ketoglutamicum*, *Brevibacterium linens*, *Brevibacterium sp.*, *Streptomyces lividans*, *Rhizobium leguminosarum* oder *Agrobacterium tumefaciens*. Vorteilhafterweise werden Bakterien verwendet, die schon eine erhöhte natürliche Biotinproduk-

tion besitzen.

Die taxonomische Stellung der aufgeführten Gattungen unterlag in den letzten Jahren einem starken Wandel und befindet sich noch
5 immer im Fluß, da falsche Gattungs- und Artnamen korrigiert werden. Aufgrund dieser in der Vergangenheit häufig erforderlichen taxonomischen Umgruppierungen der genannten Gattungen innerhalb der bakteriellen Systematik sind auch andere als die oben genannten Familien, Gattungen und Arten für das erfindungsgemäße Ver-
10 fahren geeignet.

Als eukaryontische Wirtsorganismen des erfindungsgemäßen Verfahrens kommen prinzipiell alle Biotin synthetisierenden Organismen in Frage wie Pilze, Hefen, Pflanzen oder pflanzliche Zellen. Als
15 Hefen seien die Gattungen *Rhodotorula*, *Yarrowia*, *Sporobolomyces*, *Saccharomyces* oder *Schizosaccharomyces* bevorzugt genannt. Besonders bevorzugt sind die Gattungen und Arten *Rhodotorula rubra*, *Rhodotorula glutinis*, *Rhodotorula graminis*, *Yarrowia lipolytica*, *Sporobolomyces salmonicolor*, *Sporobolomyces shibatanus* oder *Sac-*
20 *charomyces cerevisiae*.

Als Wirtsorganismus können prinzipiell alle Pflanzen verwendet werden, bevorzugt werden Pflanzen, die in der Tierernährung oder in der humanen Ernährung eine Rolle spielen wie Mais, Weizen,
25 Gerste, Roggen, Kartoffeln, Erbsen, Bohnen, Sonnenblumen, Palmen, Hirse, Sesam, Kopra oder Raps. Auch Pflanzen wie *Arabidopsis thaliana* oder *Lavendula vera* sind geeignet. Besonders bevorzugt werden pflanzliche Zellkulturen, Protoplasten aus Pflanzen oder Kalkulturen.

30 Vorteilhafterweise werden im erfindungsgemäßen Verfahren Mikroorganismen wie Bakterien, Pilze, Hefen oder pflanzliche Zellen verwendet, die in der Lage sind Biotin in das Anzuchtmedium auszuscheiden und die gegebenenfalls zusätzlich schon eine erhöhte natürliche Biotinsynthese haben. Vorteilhafterweise können diese
35 Organismen noch bezüglich der Regulation ihrer Biotinbiosynthese defekt sein, das heißt es findet keine oder nur eine sehr verringerte Regulation der Synthese statt. Dieser Regulationsdefekt hat zur Folge, daß diese Organismen schon eine wesentlich höhere Biotinproduktivität besitzen. Ein solcher Regulationsdefekt ist beispielsweise von *Escherichia coli* als *birA*-Defektmutanten bekannt
40 und sollte vorzugsweise in Form eines durch äußere Einflüsse induzierbaren Defektes, beispielsweise temperaturinduzierbar, in den Zellen vorhanden sein. Es können im Prinzip auch Organismen
45 verwendet werden, die keine natürliche Biotinproduktion aufweisen, nachdem sie mit den Biotingenen transformiert wurden.

8

Um die Biotinproduktivität insgesamt weiter zu steigern sollten die Organismen im erfindungsgemäßen Verfahren vorteilhafterweise zusätzlich mindesten ein weiteres Biotingen ausgewählt aus der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR enthalten. Vorteilhafterweise können auch solche Gene in Kombination mit der Sequenz SEQ ID No. 1, der SEQ ID NO: 3, der SEQ ID No.5 oder der SEQ ID NO.7 und ihrer Kombinationen in der Zelle vorhanden sein, die die Biotinsynthese stimulieren. Gene die die Biotinsynthese stimulieren sind z. B. das Flavore-

10 doxin-Gen, die Flavoredoxin-Reduktase-Gen. Dieses zusätzliche Gen oder diese zusätzlichen Gene können wie auch die Gene mit der Sequenz SEQ ID No. 1, der SEQ ID NO.3, SEQ ID No.5 oder der SEQ ID NO.7 oder deren Kombinationen in ein oder mehreren Kopien in der Zelle vorhanden sein. Sie können auf dem gleichen Vektor wie die

15 Sequenz SEQ ID No. 1, SEQ ID NO.3, SEQ ID No.5 und/oder SEQ ID No.7 lokalisiert sein oder auf getrennten Vektoren oder aber chromosomal integriert worden sein. Auch die Sequenzen SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 und/oder SEQ ID No.7 können zusammen auf einem Vektor oder auf getrennten Vektoren sein oder

20 ins Genom inseriert werden.

Unter dem erfindungsgemäßen Genkonstrukt sind die Gensequenzen des SAM-Synthase-Gens SEQ ID No. 1 und der Biotinsynthesegene SEQ ID No.3, SEQ ID No.5 und/oder SEQ ID No.7, sowie deren funktionelle Varianten, Analoge oder Derivate zu verstehen, die mit einem oder mehreren Regulationssignalen zur Erhöhung der Genexpression funktionell verknüpft wurden. Zusätzlich zu diesen neuen Regulationssequenzen kann die natürliche Regulation dieser Sequenzen vor den eigentlichen Strukturgenen noch vorhanden sein und gegebenenfalls genetisch verändert worden sein, so daß die natürliche Regulation ausgeschaltet und die Expression der Gene erhöht wurde. Das Genkonstrukt kann aber auch einfacher aufgebaut sein, das heißt es werden keine zusätzlichen Regulationssignale vor die Sequenzen SEQ ID No. 1, SEQ ID No. 3, SEQ ID No.5 und/oder SEQ ID

25 No.7 inseriert und der natürliche Promotor mit seiner Regulation wird nicht entfernt. Stattdessen wird die natürliche Regulationssequenz so mutiert, daß keine Regulation durch Biotin mehr erfolgt und die Genexpression gesteigert wird. Die Sequenzen SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 und/oder SEQ ID No.7 können unter der Regulation eines Promotors oder unter der Regulation getrennter Promotoren liegen. Auch am 3'-Ende der DNA-Sequenzen können

30 zusätzliche vorteilhafte regulatorische Elemente inseriert werden. Die Gene mit den Sequenzen SEQ ID No. 1, SEQ ID No. 3, SEQ ID No.5 oder SEQ ID No. 7 können in einer oder mehreren Kopien im Genkonstrukt enthalten sein.

35

40

45

Vorteilhafte Regulationssequenzen für das erfindungsgemäße Verfahren sind beispielsweise in Promotoren wie cos-, tac-, trp-,

tet-, trp-tet-, lpp-, lac-, lpp-lac-, lacI^q-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, λ -P_R- oder im λ -P_L-Promotor enthalten, die vorteilhafterweise in gram-negativen Bakterien Anwendung finden. Weitere vorteilhafte Regulationssequenzen sind beispielsweise in
5 den gram-positiven Promotoren amy und SPO2, in den Hefepromotoren ADC1, MF α , AC, P-60, CYC1, GAPDH oder in den Pflanzenpromotoren CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos oder im Ubiquitin- oder Phaseolin-Promotor enthalten.

10 Prinzipiell können alle natürlichen Promotoren mit ihren Regulationssequenzen wie die oben genannten für das erfindungsgemäße Verfahren verwendet werden. Darüberhinaus können auch synthetische Promotoren vorteilhaft verwendet werden.

15 Im Genkonstrukt können weitere Biotingene ausgewählt aus der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR in einer oder mehreren Kopien enthalten sein, die einen eigenen Promotor haben können oder aber unter der Regulation des Promotors einer der Sequenzen oder unter der Regulation des Pro-
20 motors der gesamten Sequenzen SEQ ID No. 1, SEQ ID No. 3, SEQ ID No.5 oder SEQ ID No.7 liegen können.

Das Genkonstrukt wird zur Expression in den oben genannten Wirtsorganismus vorteilhafterweise in einen wirtsspezifischen Vektor
25 inseriert, der eine optimale Expression der Gene im Wirt ermöglicht. Vektoren sind dem Fachmann wohl bekannt und können beispielsweise aus dem Buch Cloning Vectors (Eds. Pouwels P. H. et al. Elsevier, Amsterdam-New York-Oxford, 1985, ISBN 0 444 904018) entnommen werden. Unter Vektoren sind außer Plasmiden auch alle
30 anderen dem Fachmann bekannten Vektoren wie beispielsweise Phagen, Viren, Transposons, IS-Elemente, Phasmide, Cosmide, lineare oder zirkuläre DNA zu verstehen. Diese Vektoren können autonom im Wirtsorganismus repliziert oder chromosomal repliziert werden.

35 Unter Expressionssysteme sind die Kombination aus den oben beispielhaft genannten Wirtsorganismen und den zu den Organismen passenden Vektoren wie Plasmide, Viren oder Phagen wie beispielsweise Plasmide mit dem RNA-Polymerase/Promoter System, die Phagen λ , Mu oder andere temperante Phagen oder Transposons und/oder wei-
40 teren vorteilhaften regulatorischen Sequenzen zu verstehen.

Bevorzugt sind unter dem Begriff Expressionssysteme die Kombination aus Escherichia coli und seinen Plasmiden und Phagen und den dazugehörenden Promotoren sowie Bacillus und seine Plasmide und
45 Promotoren zu verstehen.

10

Für die vorteilhafte erfindungsgemäße Expression der SEQ ID No.1, SEQ ID No.3, SEQ ID No.5 und/oder SEQ ID No. 7 sind außerdem weitere 3' und/oder 5' Terminale regulatorische Sequenzen geeignet.

- 5 Diese regulatorischen Sequenzen sollen die gezielte Expression der Biotingene und der Proteinexpression ermöglichen. Dies kann beispielsweise je nach Wirtsorganismus bedeuten, daß das Gen erst nach Induktion exprimiert oder überexprimiert wird, oder daß es sofort exprimiert und/oder überexprimiert wird.
- 10

- Die regulatorischen Sequenzen bzw. Faktoren können dabei vorzugsweise die Biotingenexpression positiv beeinflussen und dadurch erhöhen. So kann eine Verstärkung der regulatorischen Elemente vorteilhafterweise auf der Transkriptionsebene erfolgen, indem
- 15 starke Transkriptionssignale wie Promotoren und/oder "Enhancer" verwendet werden. Daneben ist aber auch eine Verstärkung der Translation möglich, indem beispielsweise die Stabilität der mRNA verbessert wird.

- 20 Unter "Enhancer" sind beispielsweise DNA-Sequenzen zu verstehen, die über eine verbesserte Wechselwirkung zwischen RNA-Polymerase und DNA eine erhöhte Biotingenexpression bewirken.

- 25 Eine Steigerung der von der Sequenz SEQ ID No. 1, SEQ ID No.3, SEQ ID No.5 und SEQ ID No.7 abgeleiteten Proteinen (siehe SEQ ID No.2, SEQ ID No.4, SEQ ID No.6 und SEQ ID No.8) und ihrer Enzymaktivität läßt sich zum Beispiel gegenüber den Ausgangsenzymen durch Veränderung der entsprechenden Gensequenzen oder der Se-
- 30 quenzen seiner Homologen durch klassische Mutagenese wie UV-Be- strahlung oder Behandlung mit chemischen Mutagentien und/oder durch gezielte Mutagenese wie site directed mutagenesis, Dele- tion(en), Insertion(en) und/oder Substitution(en) erzielen. Auch kann eine erhöhte Enzymaktivität neben der beschriebenen Genam-
- 35 plifikation durch Ausschaltung von Faktoren, die die Enzybio- synthese reprimieren und/oder durch Synthese aktiver statt inak- tiver Enzyme erreicht werden.

- Durch das erfindungsgemäße Verfahren wird die Umwandlung von DTB
- 40 in Biotin und damit die Biotinproduktivität insgesamt über die in die Organismen über Vektoren und/oder chromosomal klonierten ein- gebrachten Biotingene mit der Sequenz SEQ ID No. 1, der SEQ ID No.3, SEQ ID No.5 und der SEQ ID No.7 und der Kombinationen der Gene der Sequenz SEQ ID No.1 und der SEQ ID No.5 oder SEQ ID No.1
- 45 und der SEQ ID No.7, bevorzugterweise die Kombination der Gene der Sequenz SEQ ID No.1 und der SEQ ID No.3 vorteilhaft gesteigert.

gert.

Im erfindungsgemäßen Verfahren werden die SEQ ID No.1, SEQ ID No.3, SEQ ID No.5 und/oder SEQ ID No.7 enthaltenen Mikroorganismen in einem Medium, das das Wachstum dieser Organismen ermöglicht, angezüchtet. Dieses Medium kann ein synthetisches oder ein natürliches Medium sein. Je nach Organismus werden dem Fachmann bekannte Medien verwendet. Für das Wachstum der Mikroorganismen enthalten die verwendeten Medien eine Kohlenstoffquelle, eine Stickstoffquelle, anorganische Salze und gegebenenfalls geringe Mengen an Vitamine und Spurenelemente.

Vorteilhafte Kohlenstoffquellen sind beispielsweise Zucker wie Mono-, Di- oder Polysaccharide wie Glucose, Fructose, Mannose, Xylose, Galactose, Ribose, Sorbose, Ribulose, Lactose, Maltose, Saccharose, Raffinose, Stärke oder Cellulose, komplexe Zuckerquellen wie Melasse, Zuckerphosphate wie Fructose-1,6-bisphosphat, Zuckeralkohole wie Mannit, Polyole wie Glycerin, Alkohole wie Methanol oder Ethanol, Carbonsäuren wie Citronensäure, Milchsäure oder Essigsäure, Fette wie Sojaöl oder Rapsöl, Aminosäuren wie Glutaminsäure oder Asparaginsäure oder Aminosäuren, die auch gleichzeitig als Stickstoffquelle verwendet werden können.

Vorteilhafte Stickstoffquellen sind organische oder anorganische Stickstoffverbindungen oder Materialien, die diese Verbindungen enthalten. Beispiele sind Ammoniumsalze wie NH_4Cl oder $(\text{NH}_4)_2\text{SO}_4$, Nitrate, Harnstoff, oder komplexe Stickstoffquellen wie Maisquellwasser, Bierhefeautolysat, Sojabohnenmehl, Weizengluten, Hefeextrakt, Fleischextrakt, Caseinhydrolysat, Hefe oder Kartoffelprotein, die häufig auch gleichzeitig als Stickstoffquelle dienen können.

Beispiele für anorganische Salze sind die Salze von Calcium, Magnesium, Natrium, Mangan, Kalium, Zink, Kupfer und Eisen. Als Anion dieser Salze sind besonders das Chlor-, Sulfat- und Phosphation zu nennen. Ein wichtiger Faktor zur Steigerung der Produktivität im erfindungsgemäßen Verfahren ist der Zusatz von Fe^{2+} - oder Fe^{3+} -Salzen und/oder Kaliumsalzen zum Produktionsmedium.

Gegebenenfalls werden dem Nährmedium weitere Wachstumsfaktoren zugesetzt, wie beispielsweise Vitamine oder Wachstumsförderer wie Riboflavin, Thiamin, Folsäure, Nicotinsäure, Pantothenat oder Pyridoxin, Aminosäuren wie Alanin, Cystein, Asparagin, Asparaginsäure, Glutamin, Serin, Methionin oder Lysin, Carbonsäuren wie Citronensäure, Ameisensäure, Pimelinsäure oder Milchsäure, oder Substanzen wie Dithiothreitol.

12

Zur Stabilisierung der Vektoren mit den Biotingenen in den Zellen können gegebenenfalls Antibiotika dem Medium zugesetzt werden.

Das Mischungsverhältnis der genannten Nährstoffe hängt von der Art der Fermentation ab und wird im Einzelfall festgelegt. Die Mediumkomponenten können alle zu Beginn der Fermentation vorgelegt werden, nachdem sie falls erforderlich getrennt sterilisiert oder gemeinsam sterilisiert wurden, oder aber je nach Bedarf während der Fermentation nachgegeben werden.

10

Die Züchtungsbedingungen werden so festgelegt, daß die Organismen optimal wachsen und daß die bestmöglichen Ausbeuten erreicht werden. Bevorzugte Züchtungstemperaturen liegen bei 15 °C bis 40 °C. Besonders vorteilhaft sind Temperaturen zwischen 25 °C und 37 °C.

Vorzugsweise wird der pH-Wert in einem Bereich von 3 bis 9 festgehalten. Besonders vorteilhaft sind pH-Werte zwischen 5 und 8. Im allgemeinen ist eine Inkubationsdauer von 8 bis 240 Stunden bevorzugt von 8 bis 120 Stunden ausreichend. Innerhalb dieser Zeit reichert sich die maximale Menge an Biotin im Medium an und/oder ist nach Aufschluß der Zellen verfügbar.

20

Das erfindungsgemäße Verfahren zur Herstellung von Biotin kann kontinuierlich oder batch- oder fed-batch-weise durchgeführt werden. Werden aus den mit den Biotingenen transformierten Pflanzenzellen ganze Pflanzen regeneriert, so können diese nach dem erfindungsgemäßen Verfahren ganz normal angezüchtet und vermehrt werden.

25

Beispiele:

30

1. Klonierung des S-Adenosyl-Methionin-Synthase-Gens (SEQ ID No.1):

Das Gen, das für SAM-Synthase (metK) kodiert, wurde aus dem Chromosom von E. coli durch eine Polymerase-Kettenreaktion mit Hilfe zweier spezifischer Oligonukleotide ausgehend von genomischer E. coli DNA amplifiziert. Die derart amplifizierte DNA wurde aufgereinigt, mit dem Restriktionsenzym Acc65I verdaut und in einen mit dem gleichen Enzym geschnittenen Vektor inseriert, der eine Überexpression des Gens in E. coli Stämmen ermöglicht. Durch eines der beiden Oligonukleotide wurde das Genkonstrukt mit orientierten Translationssignalen versehen

40

a.) Entwicklung von Oligonukleotiden zur Amplifizierung des metK Gens aus dem E. coli Chromosom:

45

13

metK soll als Expressionskassette bestehend aus einer ribosomalen Bindungsstelle, dem Startkodon der kodierenden Sequenz und dem Stopkodon zwischen zwei Erkennungstellen für Restriktionsenzyme amplifiziert werden. Für beide Restriktionsschnittstellen wurde die Erkennungssequenz von Acc65I gewählt. Das metK Gen wurde mit Hilfe der Oligonucleotide PmetK1 (5'- GCGGTACCAGGTGATATTAATATG-GCAAAAC-3') und PmetK2 (5'-CGGGTACCGATTACTTCAGACCGGCAGC-3') amplifiziert und kloniert.

10 b.) Durchführung der PCR:

Bedingungen:

15 Als Matrize wurden 0,5 µg chromosomale DNA von E. coli W3110 verwendet. Die Oligonukleotide PmetK1 und PmetK2 wurden in einer Konzentration von je 15 pMol eingesetzt. Die Konzentration an dNTP's betrug 200 µM. Als Polymerase wurden 2,5 U Pwo DNA-Polymerase (Boehringer Mannheim) im Reaktionspuffer des Herstellers eingesetzt. Das Volumen der PCR-Reaktion betrug 100 µl.

20

Amplifikationsbedingungen:

Die Denaturierung der DNA erfolgte für 2 min bei 94 °C. Anschließend wurden die Oligonukleotide für 30 sec bei 55 °C angelagert. 25 Die Elongation erfolgte für 75 sec bei 72 °C. Die PCR-Reaktion wurde über 30 Zyklen durchgeführt.

Das erhaltene DNA-Produkt mit einer Größe von ca. 1145 bp wurde aufgereinigt und durch Acc65I im geeigneten Puffer verdaut. 30

c.) Klonierung von metK in Expressionsvektor

2 µg des Vektors pHS1 (Konstruktion wurde in DE 197.31274.8, Priorität 22.7.97, Beispiele 1. Seite 14 bis 17 beschrieben) wurden durch Acc65I verdaut und durch Shrimp Alkalische Phosphatase (SAP) (Boehringer Mannheim) dephosphoryliert. Nach Denaturierung der SAP wurden Vektor und Fragment in einem molaren Verhältnis von 1:3 durch den Rapid-DNA-Ligation Kit nach der Vorschrift des Herstellers ligiert. Die Transformation des Ligationsansatzes erfolgte in den Stamm E. coli XL-1-blue. Positive Klone wurden durch Plasmidpräparation und Restriktionsanalyse identifiziert. Die richtige Orientierung des MetK-Fragments in pHS1 wurde durch Restriktionsverdau und Sequenzierung bestimmt. Das erhaltene Konstrukt wurde pHS1 metK (Figur 1) genannt. Die Sequenz von pHS1 metK ist SEQ ID No.9 zu entnehmen. SEQ ID No.10 zeigt die abgeleitete Aminosäuresequenz der codierenden Region für metK. 45

2. Konstruktion der Plasmide pHBbio14 und pHS1 bioS1

Die Konstruktion der Plasmide pHBbio14 und pHS1 bioS1 wurde bereits beschrieben (DE 197.31274.8, Priorität 22.7.97, Beispiele 1, 2 und 5).

3. Konstruktion von pHS1 metK bioS1

Die Plasmide pHS1 bioS1 [SEQ ID No.11, (DE 197.31274.8, Priorität 22.7.97), SEQ ID No.12 zeigt die abgeleitete Aminosäuresequenz der codierenden Region für bioS1] und pHS1 metK SEQ ID No.9 wurden durch einen Plasmid-Präparationsmethode (Boehringer) aufgereinigt. Aus pHS1 metK wurde das metK-Gen tragende Fragment durch einen Acc65I-Verdau isoliert. pHS1 bioS1 wurde durch Acc65I verdaut, und durch Shrimp Alkalische Phosphatase (SAP) (Boehringer Mannheim) dephosphoryliert. Nach Denaturierung der SAP nach Vorschrift des Herstellers wurden Vektor und das metK Fragment in einem molaren Verhältnis von 1:3 durch den Rapid-DNA-Ligation Kit nach der Vorschrift des Herstellers ligiert. Die Transformation des Ligationsansatzes erfolgte in den Stamm E. coli XL-1-blue. Positive Klone wurden durch Plasmidpräparation und Restriktionsanalyse identifiziert. Die richtige Orientierung des metK-Fragments in pHS1 bioS1 wurde durch Restriktionsverdau und Sequenzierung bestimmt. Das erhaltene Konstrukt wurde pHS1 metK bioS1 (Figur 2) genannt. Die Sequenz von pHS1 metK bioS1 ist SEQ ID No.13 zu entnehmen. SEQ ID No.14 zeigt die abgeleitete Aminosäuresequenz der codierenden Region für metK, SEQ ID No.15 zeigt die abgeleitete Aminosäuresequenz der codierenden Region für bioS1

4. Erhöhung der Biotin-Produktivität durch Überexpression von metK, bioS1 und metK in Kombination mit bioS1.

Vom Stamm BM4086 (Ketner und Campbell J. Molec. Biology 1975 96:13) wurde durch Plattieren auf Rifampycin-Platten spontan-Rifampycin-resistente Kolonien isoliert. Von einem dieser resistenten Stämme wurde ein P1-Lysat erzeugt. Mit diesem P1-Lysat wurde der Stamm W3110 transduziert und anschließend Klone durch Rifampycin selektioniert. Der erhaltene Stamm mit dem Plasmid pHBbio14 nach der CaCl₂-Methode transformiert (Maniatis et al. Molecular Cloning Cols Spring Harbour Laboratory Press 1989) und auf LB-Ampizillin 100 µg/ml angezogen. Der isolierte transformierte Stamm (LU5560), wurde jeweils mit dem Plasmiden pHS1, pHS1 metK, pHS1 bioS1 oder pHS1 metK bioS1 nach der CaCl₂-Methode transformiert und auf LB-Agar mit Ampizillin 100 µg/ml und Kanamycin 25 µg/ml selektioniert.

15

Je eine Kolonie der jeweiligen Transformanden wurde in einer DYT-Kultur mit dem entsprechenden Antibiotika angeimpft und für 12 h inkubiert. Die Übernachtskultur (= ÜNK) wurde eingesetzt um eine 10 ml Kultur in TB-Medium (Sambrook, J. Fritsch, E F. Maniatis, T.

5 2nd ed. Cold Spring Harbor Laboratory Press., 1989

ISBN 0-87969-373-8), das 30g/l Glycerol enthält, mit den entsprechenden Antibiotika anzuimpfen. Im Fall der Gegenwart der Plasmide pHS1, pHS1 metK, pHS1 bioS1 und pHS1 metK bioS1 erfolgte gleichzeitig der Zusatz von 1mM IPTG und 0,5% Arabinose zur Induktion der Genexpression von metK und bioS1 bzw. der Kombination beider
10 Gene. Nach 24h Stunden wurden die Zellen vom Kulturüberstand durch Zentrifugation abgetrennt und die Biotin-Konzentration durch einen kompetitiven ELISA mit Streptavidin im Überstand bestimmt. Die Ergebnisse dieser Bestimmung sind Tabelle I zu entnehmen.
15

Tabelle I: Bestimmung der Biotinkonzentration

	Stamm	Plasmid I	Plasmid II	Biotin mg/l
20	LU5580	pHBbio14	Kontrolle, ohne Plasmid	11
	LU5580	pHBbio14	pHS1	25
	LU5580	pHBbio14	pHS1 bioS1	45
	LU5580	pHBbio14	pHS1 metK	37
25	LU5580	pHBbio14	pHS1 metK bioS1	52

30

35

40

45

16

SEQUENZPROTOKOLL

(1) ALGEMEINE INFORMATION:

(i) ANMELDER:

- (A) NAME: BASF Aktiengesellschaft
- (B) STRASSE: Karl Bosch Strasse
- (C) ORT: Ludwigshafen
- (D) BUNDESLAND: Rheinland-Pfalz
- (E) LAND: Deutschland
- (F) POSTLEITZAHL: 67056

(ii) ANMELDETITEL: Verfahren zur Herstellung von Biotin

(iii) ANZAHL DER SEQUENZEN: 15

(iv) COMPUTER-LESBARE FORM:

- (A) DATENTRÄGER: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPA)

(2) INFORMATION ZU SEQ ID NO: 1:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 1155 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

(vi) URSPRÜNGLICHE HERKUNFT:

- (B) STAMM: Escherichia coli

(vii) UNMITTELBARE HERKUNFT:

- (B) CLON: metK

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: CDS
- (B) LAGE: 1..1155

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

ATG GCA AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT
Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro

1

5

10

15

48

17

GAC	AAA	ATT	GCT	GAC	CAA	ATT	TCT	GAT	GCC	GTT	TTA	GAC	GCG	ATC	CTC	96
Asp	Lys	Ile	Ala	Asp	Gln	Ile	Ser	Asp	Ala	Val	Leu	Asp	Ala	Ile	Leu	
			20						25				30			
GAA	CAG	GAT	CCG	AAA	GCA	CGC	GTT	GCT	TGC	GAA	ACC	TAC	GTA	AAA	ACC	144
Glu	Gln	Asp	Pro	Lys	Ala	Arg	Val	Ala	Cys	Glu	Thr	Tyr	Val	Lys	Thr	
			35					40				45				
GGC	ATG	GTT	TTA	GTT	GGC	GGC	GAA	ATC	ACC	ACC	AGC	GCC	TGG	GTA	GAC	192
Gly	Met	Val	Leu	Val	Gly	Gly	Glu	Ile	Thr	Thr	Ser	Ala	Trp	Val	Asp	
			50				55				60					
ATC	GAA	GAG	ATC	ACC	CGT	AAC	ACC	GTT	CGC	GAA	ATT	GGC	TAT	GTG	CAT	240
Ile	Glu	Glu	Ile	Thr	Arg	Asn	Thr	Val	Arg	Glu	Ile	Gly	Tyr	Val	His	
			65			70				75					80	
TCC	GAC	ATG	GGC	TTT	GAC	GCT	AAC	TCC	TGT	GCG	GTT	CTG	AGC	GCT	ATC	288
Ser	Asp	Met	Gly	Phe	Asp	Ala	Asn	Ser	Cys	Ala	Val	Leu	Ser	Ala	Ile	
				85					90					95		
GGC	AAA	CAG	TCT	CCT	GAC	ATC	AAC	CAG	GGC	GTT	GAC	CGT	GCC	GAT	CCG	336
Gly	Lys	Gln	Ser	Pro	Asp	Ile	Asn	Gln	Gly	Val	Asp	Arg	Ala	Asp	Pro	
				100				105					110			
CTG	GAA	CAG	GGC	GCG	GGT	GAC	CAG	GGT	CTG	ATG	TTT	GGC	TAC	GCA	ACT	384
Leu	Glu	Gln	Gly	Ala	Gly	Asp	Gln	Gly	Leu	Met	Phe	Gly	Tyr	Ala	Thr	
			115				120					125				
AAT	GAA	ACC	GAC	GTG	CTG	ATG	CCA	GCA	CCT	ATC	ACC	TAT	GCA	CAC	CGT	432
Asn	Glu	Thr	Asp	Val	Leu	Met	Pro	Ala	Pro	Ile	Thr	Tyr	Ala	His	Arg	
			130				135				140					
CTG	GTA	CAG	CGT	CAG	GCT	GAA	GTG	CGT	AAA	AAC	GGC	ACT	CTG	CCG	TGG	480
Leu	Val	Gln	Arg	Gln	Ala	Glu	Val	Arg	Lys	Asn	Gly	Thr	Leu	Pro	Trp	
			145			150				155					160	
CTG	CGC	CCG	GAC	GCG	AAA	AGC	CAG	GTG	ACT	TTT	CAG	TAT	GAC	GAC	GGC	528
Leu	Arg	Pro	Asp	Ala	Lys	Ser	Gln	Val	Thr	Phe	Gln	Tyr	Asp	Asp	Gly	
				165					170				175			
AAA	ATC	GTT	GGT	ATC	GAT	GCT	GTC	GTG	CTT	TCC	ACT	CAG	CAC	TCT	GAA	576
Lys	Ile	Val	Gly	Ile	Asp	Ala	Val	Val	Leu	Ser	Thr	Gln	His	Ser	Glu	
			180					185					190			
GAG	ATC	GAC	CAG	AAA	TCG	CTG	CAA	GAA	GCG	GTA	ATG	GAA	GAG	ATC	ATC	624
Glu	Ile	Asp	Gln	Lys	Ser	Leu	Gln	Glu	Ala	Val	Met	Glu	Glu	Ile	Ile	
			195				200					205				
AAG	CCA	ATT	CTG	CCC	GCT	GAA	TGG	CTG	ACT	TCT	GCC	ACC	AAA	TTC	TTC	672
Lys	Pro	Ile	Leu	Pro	Ala	Glu	Trp	Leu	Thr	Ser	Ala	Thr	Lys	Phe	Phe	
			210				215				220					

18

[illegible]

(2) INFORMATION ZU SEQ ID NO: 2:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 384 Aminosäuren

(B) ART: Aminosäure

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

19

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Met	Ala	Lys	His	Leu	Phe	Thr	Ser	Glu	Ser	Val	Ser	Glu	Gly	His	Pro	
1				5					10					15		
Asp	Lys	Ile	Ala	Asp	Gln	Ile	Ser	Asp	Ala	Val	Leu	Asp	Ala	Ile	Leu	
			20					25					30			
Glu	Gln	Asp	Pro	Lys	Ala	Arg	Val	Ala	Cys	Glu	Thr	Tyr	Val	Lys	Thr	
		35					40					45				
Gly	Met	Val	Leu	Val	Gly	Gly	Glu	Ile	Thr	Thr	Ser	Ala	Trp	Val	Asp	
	50					55					60					
Ile	Glu	Glu	Ile	Thr	Arg	Asn	Thr	Val	Arg	Glu	Ile	Gly	Tyr	Val	His	
	65				70					75					80	
Ser	Asp	Met	Gly	Phe	Asp	Ala	Asn	Ser	Cys	Ala	Val	Leu	Ser	Ala	Ile	
				85					90					95		
Gly	Lys	Gln	Ser	Pro	Asp	Ile	Asn	Gln	Gly	Val	Asp	Arg	Ala	Asp	Pro	
			100					105					110			
Leu	Glu	Gln	Gly	Ala	Gly	Asp	Gln	Gly	Leu	Met	Phe	Gly	Tyr	Ala	Thr	
		115					120					125				
Asn	Glu	Thr	Asp	Val	Leu	Met	Pro	Ala	Pro	Ile	Thr	Tyr	Ala	His	Arg	
	130					135					140					
Leu	Val	Gln	Arg	Gln	Ala	Glu	Val	Arg	Lys	Asn	Gly	Thr	Leu	Pro	Trp	
	145				150					155					160	
Leu	Arg	Pro	Asp	Ala	Lys	Ser	Gln	Val	Thr	Phe	Gln	Tyr	Asp	Asp	Gly	
				165					170					175		
Lys	Ile	Val	Gly	Ile	Asp	Ala	Val	Val	Leu	Ser	Thr	Gln	His	Ser	Glu	
			180					185					190			
Glu	Ile	Asp	Gln	Lys	Ser	Leu	Gln	Glu	Ala	Val	Met	Glu	Glu	Ile	Ile	
		195					200					205				
Lys	Pro	Ile	Leu	Pro	Ala	Glu	Trp	Leu	Thr	Ser	Ala	Thr	Lys	Phe	Phe	
	210					215					220					
Ile	Asn	Pro	Thr	Gly	Arg	Phe	Val	Ile	Gly	Gly	Pro	Met	Gly	Asp	Cys	
	225				230					235					240	
Gly	Leu	Thr	Gly	Arg	Lys	Ile	Ile	Val	Asp	Thr	Tyr	Gly	Gly	Met	Ala	
				245					250					255		
Arg	His	Gly	Gly	Gly	Ala	Phe	Ser	Gly	Lys	Asp	Pro	Ser	Lys	Val	Asp	
			260					265					270			

20

Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
 275 280 285
 Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
 290 295 300
 Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
 305 310 315 320
 Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
 325 330 335
 Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
 340 345 350
 Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
 355 360 365
 Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
 370 375 380

(2) INFORMATION ZU SEQ ID NO: 3:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 1206 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

(vi) URSPRÜNGLICHE HERKUNFT:

- (B) STAMM: Escherichia coli

(vii) UNMITTELBARE HERKUNFT:

- (B) CLON: bios1

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: CDS
- (B) LÄGE: 1..1206

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

ATG AAC GTT TTT AAT CCC GCG CAG TTT CGC GCC CAG TTT CCC GCA CTA
 Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
 1 5 10 15 48
 CAG GAT GCG GGC GTC TAT CTC GAC AGC GCC GCG ACC GCG CTT AAA CCT
 Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
 20 25 30 96

																	21		
GAA	GCC	GTG	GTT	GAA	GCC	ACC	CAA	CAG	TTT	TAC	AGT	CTG	AGC	GCC	GGA			144	
Glu	Ala	Val	Val	Glu	Ala	Thr	Gln	Gln	Phe	Tyr	Ser	Leu	Ser	Ala	Gly				
		35					40					45							
AAC	GTC	CAT	CGC	AGC	CAG	TTT	GCC	GAA	GCC	CAA	CGC	CTG	ACC	GCG	CGT			192	
Asn	Val	His	Arg	Ser	Gln	Phe	Ala	Glu	Ala	Gln	Arg	Leu	Thr	Ala	Arg				
	50					55					60								
TAT	GAA	GCT	GCA	CGA	GAG	AAA	GTG	GCG	CAA	TTA	CTG	AAT	GCA	CCG	GAT			240	
Tyr	Glu	Ala	Ala	Arg	Glu	Lys	Val	Ala	Gln	Leu	Leu	Asn	Ala	Pro	Asp				
	65				70					75					80				
GAT	AAA	ACT	ATC	GTC	TGG	ACG	CGC	GGC	ACC	ACT	GAA	TCC	ATC	AAC	ATG			288	
Asp	Lys	Thr	Ile	Val	Trp	Thr	Arg	Gly	Thr	Thr	Glu	Ser	Ile	Asn	Met				
			85						90					95					
GTG	GCA	CAA	TGC	TAT	GCG	CGT	CCG	CGT	CTG	CAA	CCG	GGC	GAT	GAG	ATT			336	
Val	Ala	Gln	Cys	Tyr	Ala	Arg	Pro	Arg	Leu	Gln	Pro	Gly	Asp	Glu	Ile				
		100						105					110						
ATT	GTC	AGC	GTG	GCA	GAA	CAC	CAC	GCC	AAC	CTC	GTC	CCC	TGG	CTG	ATG			384	
Ile	Val	Ser	Val	Ala	Glu	His	His	Ala	Asn	Leu	Val	Pro	Trp	Leu	Met				
		115					120					125							
GTC	GCC	CAA	CAA	ACT	GGA	GCC	AAA	GTG	GTG	AAA	TTG	CCG	CTT	AAT	GCG			432	
Val	Ala	Gln	Gln	Thr	Gly	Ala	Lys	Val	Val	Lys	Leu	Pro	Leu	Asn	Ala				
	130					135					140								
CAG	CGA	CTG	CCG	GAT	GTC	GAT	TTG	TTG	CCA	GAA	CTG	ATT	ACT	CCC	CGT			480	
Gln	Arg	Leu	Pro	Asp	Val	Asp	Leu	Leu	Pro	Glu	Leu	Ile	Thr	Pro	Arg				
	145				150					155					160				
AGT	CGG	ATT	CTG	GCG	TTG	GGT	CAG	ATG	TCG	AAC	GTT	ACT	GGC	GGT	TGC			528	
Ser	Arg	Ile	Leu	Ala	Leu	Gly	Gln	Met	Ser	Asn	Val	Thr	Gly	Gly	Cys				
			165					170					175						
CCG	GAT	CTG	GCG	CGA	GCG	ATT	ACC	TTT	GCT	CAT	TCA	GCC	GGG	ATG	GTG			576	
Pro	Asp	Leu	Ala	Arg	Ala	Ile	Thr	Phe	Ala	His	Ser	Ala	Gly	Met	Val				
		180						185					190						
GTG	ATG	GTT	GAT	GGT	GCT	CAG	GGG	GCA	GTG	CAT	TTC	CCC	GCG	GAT	GTT			624	
Val	Met	Val	Asp	Gly	Ala	Gln	Gly	Ala	Val	His	Phe	Pro	Ala	Asp	Val				
	195					200						205							
CAG	CAA	CTG	GAT	ATT	GAT	TTC	TAT	GCT	TTT	TCA	GGT	CAC	AAA	CTG	TAT			672	
Gln	Gln	Leu	Asp	Ile	Asp	Phe	Tyr	Ala	Phe	Ser	Gly	His	Lys	Leu	Tyr				
	210					215					220								
GGC	CCG	ACA	GGT	ATC	GGC	GTG	CTG	TAT	GGT	AAA	TCA	GAA	CTG	CTG	GAG			720	
Gly	Pro	Thr	Gly	Ile	Gly	Val	Leu	Tyr	Gly	Lys	Ser	Glu	Leu	Leu	Glu				
	225				230					235					240				

22																
GCG	ATG	TCG	CCC	TGG	CTG	GGC	GGC	GGC	AAA	ATG	GTT	CAC	GAA	GTG	AGT	768
Ala	Met	Ser	Pro	Trp	Leu	Gly	Gly	Gly	Lys	Met	Val	His	Glu	Val	Ser	
				245					250					255		
TTT	GAC	GGC	TTC	ACG	ACT	CAA	TCT	GCG	CCG	TGG	AAA	CTG	GAA	GCT	GGA	816
Phe	Asp	Gly	Phe	Thr	Thr	Gln	Ser	Ala	Pro	Trp	Lys	Leu	Glu	Ala	Gly	
			260					265					270			
ACG	CCA	AAT	GTC	GCT	GGT	GTC	ATA	GGA	TTA	AGC	GCG	GCG	CTG	GAA	TGG	864
Thr	Pro	Asn	Val	Ala	Gly	Val	Ile	Gly	Leu	Ser	Ala	Ala	Leu	Glu	Trp	
		275					280					285				
CTG	GCA	GAT	TAC	GAT	ATC	AAC	CAG	GCC	GAA	AGC	TGG	AGC	CGT	AGC	TTA	912
Leu	Ala	Asp	Tyr	Asp	Ile	Asn	Gln	Ala	Glu	Ser	Trp	Ser	Arg	Ser	Leu	
	290					295					300					
GCA	ACG	CTG	GCG	GAA	GAT	GCG	CTG	GCG	AAA	CGT	CCC	GGC	TTT	CGT	TCA	960
Ala	Thr	Leu	Ala	Glu	Asp	Ala	Leu	Ala	Lys	Arg	Pro	Gly	Phe	Arg	Ser	
305				310					315					320		
TTC	CGC	TGC	CAG	GAT	TCC	AGC	CTG	CTG	GCC	TTT	GAT	TTT	GCT	GGC	GTT	1008
Phe	Arg	Cys	Gln	Asp	Ser	Ser	Leu	Leu	Ala	Phe	Asp	Phe	Ala	Gly	Val	
			325						330					335		
CAT	CAT	AGC	GAT	ATG	GTG	ACG	CTG	CTG	GCG	GAG	TAC	GGT	ATT	GCC	CTG	1056
His	His	Ser	Asp	Met	Val	Thr	Leu	Leu	Ala	Glu	Tyr	Gly	Ile	Ala	Leu	
		340					345						350			
CGG	GCC	GGG	CAG	CAT	TGC	GCT	CAG	CCG	CTA	CTG	GCA	GAA	TTA	GGC	GTA	1104
Arg	Ala	Gly	Gln	His	Cys	Ala	Gln	Pro	Leu	Leu	Ala	Glu	Leu	Gly	Val	
		355					360					365				
ACC	GGC	ACA	CTG	CGC	GCC	TCT	TTT	GCG	CCA	TAT	AAT	ACA	AAG	AGT	GAT	1152
Thr	Gly	Thr	Leu	Arg	Ala	Ser	Phe	Ala	Pro	Tyr	Asn	Thr	Lys	Ser	Asp	
	370					375					380					
GTG	GAT	GCG	CTG	GTG	AAT	GCC	GTT	GAC	CGC	GCG	CTG	GAA	TTA	TTG	GTG	1200
Val	Asp	Ala	Leu	Val	Asn	Ala	Val	Asp	Arg	Ala	Leu	Glu	Leu	Leu	Val	
	385				390				395					400		
GAT	TAA															1206
Asp																

(2) INFORMATION ZU SEQ ID NO: 4:

- (i) SEQUENZ CHARAKTERISTIKA:
- (A) LÄNGE: 401 Aminosäuren
 - (B) ART: Aminosäure
 - (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

23

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
1 5 10 15
Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
20 25 30
Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
35 40 45
Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
50 55 60
Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
65 70 75 80
Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
85 90 95
Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
100 105 110
Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
115 120 125
Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
130 135 140
Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145 150 155 160
Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
165 170 175
Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
180 185 190
Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
195 200 205
Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
210 215 220
Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240
Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255
Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

24
Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275. 280 285
Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300
Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320
Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335
His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350
Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365
Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
370 375 380
Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385 390 395 400
Asp

(2) INFORMATION ZU SEQ ID NO: 5:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 1215 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

(vi) URSPRÜNGLICHE HERKUNFT:

- (B) STAMM: Escherichia coli

(vii) UNMITTELBARE HERKUNFT:

- (B) CLON: bios2

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: CDS
- (B) LAGE: 1..1215

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 5:

25																
ATG	AAA	TTA	CCG	ATT	TAT	CTC	GAC	TAC	TCC	GCA	ACC	ACG	CCG	GTG	GAC	48
Met	Lys	Leu	Pro	Ile	Tyr	Leu	Asp	Tyr	Ser	Ala	Thr	Thr	Pro	Val	Asp	
1				5				10						15		
96																
CCG	CGT	GTT	GCC	GAG	AAA	ATG	ATG	CAG	TTT	ATG	ACG	ATG	GAC	GGA	ACC	
Pro	Arg	Val	Ala	Glu	Lys	Met	Met	Gln	Phe	Met	Thr	Met	Asp	Gly	Thr	
			20					25					30			
144																
TTT	GGT	AAC	CCG	GCC	TCC	CGT	TCT	CAC	CGT	TTC	GGC	TGG	CAG	GCT	GAA	
Phe	Gly	Asn	Pro	Ala	Ser	Arg	Ser	His	Arg	Phe	Gly	Trp	Gln	Ala	Glu	
		35					40					45				
192																
GAA	GCG	GTA	GAT	ATC	GCC	CGT	AAT	CAG	ATT	GCC	GAT	CTG	GTC	GGC	GCT	
Glu	Ala	Val	Asp	Ile	Ala	Arg	Asn	Gln	Ile	Ala	Asp	Leu	Val	Gly	Ala	
	50					55					60					
240																
GAT	CCG	CGT	GAA	ATC	GTC	TTT	ACC	TCT	GGT	GCA	ACC	GAA	TCT	GAC	AAC	
Asp	Pro	Arg	Glu	Ile	Val	Phe	Thr	Ser	Gly	Ala	Thr	Glu	Ser	Asp	Asn	
65					70					75				80		
288																
CTG	GCG	ATC	AAA	GGT	GCA	GCC	AAC	TTT	TAT	CAG	AAA	AAA	GGC	AAG	CAC	
Leu	Ala	Ile	Lys	Gly	Ala	Ala	Asn	Phe	Tyr	Gln	Lys	Lys	Gly	Lys	His	
			85					90						95		
336																
ATC	ATC	ACC	AGC	AAA	ACC	GAA	CAC	AAA	GCG	GTA	CTG	GAT	ACC	TGC	CGT	
Ile	Ile	Thr	Ser	Lys	Thr	Glu	His	Lys	Ala	Val	Leu	Asp	Thr	Cys	Arg	
			100					105					110			
384																
CAG	CTG	GAG	CGC	GAA	GGT	TTT	GAA	GTC	ACC	TAC	CTG	GCA	CCG	CAG	CGT	
Gln	Leu	Glu	Arg	Glu	Gly	Phe	Glu	Val	Thr	Tyr	Leu	Ala	Pro	Gln	Arg	
		115					120					125				
432																
AAC	GGC	ATT	ATC	GAC	CTG	AAA	GAA	CTT	GAA	GCA	GCG	ATG	CGT	GAC	GAC	
Asn	Gly	Ile	Ile	Asp	Leu	Lys	Glu	Leu	Glu	Ala	Ala	Met	Arg	Asp	Asp	
	130					135					140					
480																
ACC	ATC	CTC	GTG	TCC	ATC	ATG	CAC	GTA	AAT	AAC	GAA	ATC	GGC	GTG	GTG	
Thr	Ile	Leu	Val	Ser	Ile	Met	His	Val	Asn	Asn	Glu	Ile	Gly	Val	Val	
145					150				155					160		
528																
CAG	GAT	ATC	GCG	GCT	ATC	GGC	GAA	ATG	TGC	CGT	GCT	CGT	GGC	ATT	ATC	
Gln	Asp	Ile	Ala	Ala	Ile	Gly	Glu	Met	Cys	Arg	Ala	Arg	Gly	Ile	Ile	
			165					170					175			
576																
TAT	CAC	GTT	GAT	GCA	ACC	CAG	AGC	GTG	GGT	AAA	CTG	CCT	ATC	GAC	CTG	
Tyr	His	Val	Asp	Ala	Thr	Gln	Ser	Val	Gly	Lys	Leu	Pro	Ile	Asp	Leu	
			180					185					190			
624																
AGC	CAG	TTG	AAA	GTT	GAC	CTG	ATG	TCT	TTC	TCC	GGT	CAC	AAA	ATC	TAT	
Ser	Gln	Leu	Lys	Val	Asp	Leu	Met	Ser	Phe	Ser	Gly	His	Lys	Ile	Tyr	
		195					200					205				

26

GGC CCG AAA GGT ATC GGT GCG CTG TAT GTA CGT CGT AAA CCG CGC GTA	672
Gly Pro Lys Gly Ile Gly Ala Leu Tyr Val Arg Arg Lys Pro Arg Val	
210 215 220	
CGC ATC GAA GCG CAA ATG CAC GGC GGC GGT CAC GAG CGC GGT ATG CGT	720
Arg Ile Glu Ala Gln Met His Gly Gly Gly His Glu Arg Gly Met Arg	
225 230 235 240	
TCC GGC ACT CTG CCT GTT CAC CAG ATC GTC GGA ATG GGC GAG GCC TAT	768
Ser Gly Thr Leu Pro Val His Gln Ile Val Gly Met Gly Glu Ala Tyr	
245 250 255	
CGC ATC GCA AAA GAA GAG ATG GCG ACC GAG ATG GAA CGT CTG CGC GGC	816
Arg Ile Ala Lys Glu Glu Met Ala Thr Glu Met Glu Arg Leu Arg Gly	
260 265 270	
CTG CGT AAC CGT CTG TGG AAC GGC ATC AAA GAT ATC GAA GAA GTT TAC	864
Leu Arg Asn Arg Leu Trp Asn Gly Ile Lys Asp Ile Glu Glu Val Tyr	
275 280 285	
CTG AAC GGT GAC CTG GAA CAC GGT GCG CCG AAC ATT CTC AAC GTC AGC	912
Leu Asn Gly Asp Leu Glu His Gly Ala Pro Asn Ile Leu Asn Val Ser	
290 295 300	
TTC AAC TAC GTT GAA GGT GAG TCG CTG ATT ATG GCG CTG AAA GAC CTC	960
Phe Asn Tyr Val Glu Gly Glu Ser Leu Ile Met Ala Leu Lys Asp Leu	
305 310 315 320	
GCA GTT TCT TCA GGT TCC GCC TGT ACG TCA GCA AGC CTC GAA CCG TCC	1008
Ala Val Ser Ser Gly Ser Ala Cys Thr Ser Ala Ser Leu Glu Pro Ser	
325 330 335	
TAC GTG CTG CGC GCG CTG GGG CTG AAC GAC GAG CTG GCA CAT AGC TCT	1056
Tyr Val Leu Arg Ala Leu Gly Leu Asn Asp Glu Leu Ala His Ser Ser	
340 345 350	
ATC CGT TTC TCT TTA GGT CGT TTT ACT ACT GAA GAA GAG ATC GAC TAC	1104
Ile Arg Phe Ser Leu Gly Arg Phe Thr Thr Glu Glu Glu Ile Asp Tyr	
355 360 365	
ACC ATC GAG TTA GTT CGT AAA TCC ATC GGT CGT CTG CGT GAC CTT TCT	1152
Thr Ile Glu Leu Val Arg Lys Ser Ile Gly Arg Leu Arg Asp Leu Ser	
370 375 380	
CCG CTG TGG GAA ATG TAC AAG CAG GGC GTG GAT CTG AAC AGC ATC GAA	1200
Pro Leu Trp Glu Met Tyr Lys Gln Gly Val Asp Leu Asn Ser Ile Glu	
385 390 395 400	
TGG GCT CAT CAT TAA	1215
Trp Ala His His	
405	

(2) INFORMATION ZU SEQ ID NO: 6:

27

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 404 Aminosäuren

(B) ART: Aminosäure

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 6:

Met	Lys	Leu	Pro	Ile	Tyr	Leu	Asp	Tyr	Ser	Ala	Thr	Thr	Pro	Val	Asp	
1				5					10						15	
Pro	Arg	Val	Ala	Glu	Lys	Met	Met	Gln	Phe	Met	Thr	Met	Asp	Gly	Thr	
		20						25					30			
Phe	Gly	Asn	Pro	Ala	Ser	Arg	Ser	His	Arg	Phe	Gly	Trp	Gln	Ala	Glu	
		35					40					45				
Glu	Ala	Val	Asp	Ile	Ala	Arg	Asn	Gln	Ile	Ala	Asp	Leu	Val	Gly	Ala	
		50				55					60					
Asp	Pro	Arg	Glu	Ile	Val	Phe	Thr	Ser	Gly	Ala	Thr	Glu	Ser	Asp	Asn	
65				70					75						80	
Leu	Ala	Ile	Lys	Gly	Ala	Ala	Asn	Phe	Tyr	Gln	Lys	Lys	Gly	Lys	His	
			85					90						95		
Ile	Ile	Thr	Ser	Lys	Thr	Glu	His	Lys	Ala	Val	Leu	Asp	Thr	Cys	Arg	
		100						105					110			
Gln	Leu	Glu	Arg	Glu	Gly	Phe	Glu	Val	Thr	Tyr	Leu	Ala	Pro	Gln	Arg	
		115					120					125				
Asn	Gly	Ile	Ile	Asp	Leu	Lys	Glu	Leu	Glu	Ala	Ala	Met	Arg	Asp	Asp	
		130				135						140				
Thr	Ile	Leu	Val	Ser	Ile	Met	His	Val	Asn	Asn	Glu	Ile	Gly	Val	Val	
145					150				155						160	
Gln	Asp	Ile	Ala	Ala	Ile	Gly	Glu	Met	Cys	Arg	Ala	Arg	Gly	Ile	Ile	
			165					170					175			
Tyr	His	Val	Asp	Ala	Thr	Gln	Ser	Val	Gly	Lys	Leu	Pro	Ile	Asp	Leu	
		180					185						190			
Ser	Gln	Leu	Lys	Val	Asp	Leu	Met	Ser	Phe	Ser	Gly	His	Lys	Ile	Tyr	
		195				200						205				
Gly	Pro	Lys	Gly	Ile	Gly	Ala	Leu	Tyr	Val	Arg	Arg	Lys	Pro	Arg	Val	
		210				215					220					
Arg	Ile	Glu	Ala	Gln	Met	His	Gly	Gly	Gly	His	Glu	Arg	Gly	Met	Arg	
225					230				235						240	

28

Ser Gly Thr Leu Pro Val His Gln Ile Val Gly Met Gly Glu Ala Tyr
245 250 255

Arg Ile Ala Lys Glu Glu Met Ala Thr Glu Met Glu Arg Leu Arg Gly
260 265 270

Leu Arg Asn Arg Leu Trp Asn Gly Ile Lys Asp Ile Glu Glu Val Tyr
275 280 285

Leu Asn Gly Asp Leu Glu His Gly Ala Pro Asn Ile Leu Asn Val Ser
290 295 300

Phe Asn Tyr Val Glu Gly Glu Ser Leu Ile Met Ala Leu Lys Asp Leu
305 310 315 320

Ala Val Ser Ser Gly Ser Ala Cys Thr Ser Ala Ser Leu Glu Pro Ser
325 330 335

Tyr Val Leu Arg Ala Leu Gly Leu Asn Asp Glu Leu Ala His Ser Ser
340 345 350

Ile Arg Phe Ser Leu Gly Arg Phe Thr Thr Glu Glu Glu Ile Asp Tyr
355 360 365

Thr Ile Glu Leu Val Arg Lys Ser Ile Gly Arg Leu Arg Asp Leu Ser
370 375 380

Pro Leu Trp Glu Met Tyr Lys Gln Gly Val Asp Leu Asn Ser Ile Glu
385 390 395 400

Trp Ala His His

(2) INFORMATION ZU SEQ ID NO: 7:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 1221 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

(vi) URSPRÜNGLICHE HERKUNFT:

- (B) STAMM: Escherichia coli

(vii) UNMITTELBARE HERKUNFT:

- (B) CLON: bioS3

29

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: CDS

(B) LAGE: 1..1221

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 7:

ATG	ATT	TTT	TCC	GTC	GAC	AAA	GTG	CGG	GCC	GAC	TTT	CCG	GTG	CTT	TCG	48
Met	Ile	Phe	Ser	Val	Asp	Lys	Val	Arg	Ala	Asp	Phe	Pro	Val	Leu	Ser	
1				5					10					15		
CGT	GAG	GTA	AAC	GGT	TTG	CCG	CTG	GCT	TAT	CTC	GAC	AGC	GCC	GCC	AGT	96
Arg	Glu	Val	Asn	Gly	Leu	Pro	Leu	Ala	Tyr	Leu	Asp	Ser	Ala	Ala	Ser	
			20					25					30			
GCG	CAG	AAA	CCG	AGC	CAG	GTG	ATT	GAC	GCC	GAG	GCC	GAG	TTT	TAT	CGT	144
Ala	Gln	Lys	Pro	Ser	Gln	Val	Ile	Asp	Ala	Glu	Ala	Glu	Phe	Tyr	Arg	
		35					40					45				
CAT	GGC	TAC	GCG	GCG	GTG	CAT	CGT	GGT	ATT	CAT	ACC	TTA	AGC	GCC	CAG	192
His	Gly	Tyr	Ala	Ala	Val	His	Arg	Gly	Ile	His	Thr	Leu	Ser	Ala	Gln	
	50					55				60						
GCG	ACC	GAG	AAA	ATG	GAG	AAC	GTG	CGC	AAG	CGG	GCA	TCG	CTG	TTT	ATT	240
Ala	Thr	Glu	Lys	Met	Glu	Asn	Val	Arg	Lys	Arg	Ala	Ser	Leu	Phe	Ile	
65					70				75						80	
AAT	GCC	CGT	TCG	GCG	GAA	GAG	CTG	GTG	TTC	GTC	CGC	GGC	ACG	ACG	GAA	288
Asn	Ala	Arg	Ser	Ala	Glu	Glu	Leu	Val	Phe	Val	Arg	Gly	Thr	Thr	Glu	
				85					90					95		
GGG	ATC	AAT	CTG	GTC	GCC	AAT	AGC	TGG	GGC	AAC	AGC	AAC	GTG	CGG	GCG	336
Gly	Ile	Asn	Leu	Val	Ala	Asn	Ser	Trp	Gly	Asn	Ser	Asn	Val	Arg	Ala	
			100					105					110			
GGC	GAT	AAC	ATC	ATC	ATC	AGT	CAG	ATG	GAG	CAC	CAC	GCT	AAC	ATT	GTT	384
Gly	Asp	Asn	Ile	Ile	Ile	Ser	Gln	Met	Glu	His	His	Ala	Asn	Ile	Val	
		115					120					125				
CCC	TGG	CAG	ATG	CTT	TGC	GCA	CGC	GTT	GGC	GCA	GAG	CTG	CGT	GTG	ATC	432
Pro	Trp	Gln	Met	Leu	Cys	Ala	Arg	Val	Gly	Ala	Glu	Leu	Arg	Val	Ile	
	130					135					140					
CCG	CTC	AAT	CCC	GAT	GGT	ACG	TTG	CAA	CTG	GAG	ACG	CTG	CCT	ACG	CTG	480
Pro	Leu	Asn	Pro	Asp	Gly	Thr	Leu	Gln	Leu	Glu	Thr	Leu	Pro	Thr	Leu	
145					150					155					160	
TTT	GAT	GAG	AAA	ACT	CGC	CTG	CTG	GCA	ATT	ACT	CAT	GTC	TCC	AAC	GTG	528
Phe	Asp	Glu	Lys	Thr	Arg	Leu	Leu	Ala	Ile	Thr	His	Val	Ser	Asn	Val	
				165					170					175		
CTT	GGC	ACA	GAA	AAT	CCA	CTG	GCG	GAA	ATG	ATC	ACG	CTT	GCG	CAC	CAG	576
Leu	Gly	Thr	Glu	Asn	Pro	Leu	Ala	Glu	Met	Ile	Thr	Leu	Ala	His	Gln	
				180				185					190			

30

CAT GGC GCA AAA GTG CTG GTG GAT GGC GCT CAG GCG GTG ATG CAT CAT	624
His Gly Ala Lys Val Leu Val Asp Gly Ala Gln Ala Val Met His His	
195 200 205	
CCG GTG GAT GTT CAG GCG CTG GAT TGC GAC TTT TAC GTG TTC TCC GGG	672
Pro Val Asp Val Gln Ala Leu Asp Cys Asp Phe Tyr Val Phe Ser Gly	
210 215 220	
CAT AAA CTG TAT GGC CCC ACC GGA ATT GGC ATT CTT TAT GTG AAA GAA	720
His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Ile Leu Tyr Val Lys Glu	
225 230 235 240	
GCC TTG TTG CAG GAG ATG CCG CCG TGG GAA GGG GGC GGT TCT ATG ATC	768
Ala Leu Leu Gln Glu Met Pro Pro Trp Glu Gly Gly Gly Ser Met Ile	
245 250 255	
GCC ACC GTC AGC CTG AGT GAA GGC ACT ACC TGG ACC AAA GCA CCA TGG	816
Ala Thr Val Ser Leu Ser Glu Gly Thr Thr Trp Thr Lys Ala Pro Trp	
260 265 270	
CGG TTT GAA GCC GGT ACA CCC AAT ACC GGG GGC ATC ATT GGT CTT GGC	864
Arg Phe Glu Ala Gly Thr Pro Asn Thr Gly Gly Ile Ile Gly Leu Gly	
275 280 285	
GCG GCG CTG GAG TAT GTT TCG GCG CTG GGG CTT AAT AAC ATA GCC GAG	912
Ala Ala Leu Glu Tyr Val Ser Ala Leu Gly Leu Asn Asn Ile Ala Glu	
290 295 300	
TAT GAA CAG AAT CTG ATG CAT TAT GCG CTA TCA CAG CTG GAA TCT GTA	960
Tyr Glu Gln Asn Leu Met His Tyr Ala Leu Ser Gln Leu Glu Ser Val	
305 310 315 320	
CCG GAT CTC ACT CTC TAT GGC CCA CAA AAC AGG CTT GGC GTT ATT GCT	1008
Pro Asp Leu Thr Leu Tyr Gly Pro Gln Asn Arg Leu Gly Val Ile Ala	
325 330 335	
TTT AAT CTC GGT AAA CAC CAC GCC TAT GAT GTT GGC AGT TTT CTC GAT	1056
Phe Asn Leu Gly Lys His His Ala Tyr Asp Val Gly Ser Phe Leu Asp	
340 345 350	
AAT TAC GGC ATT GCT GTG CGT ACC GGA CAT CAC TGC GCA ATG CCA TTG	1104
Asn Tyr Gly Ile Ala Val Arg Thr Gly His His Cys Ala Met Pro Leu	
355 360 365	
ATG GCC TAT TAC AAC GTC CCT GCG ATG TGT CGG GCG TCG CTG GCC ATG	1152
Met Ala Tyr Tyr Asn Val Pro Ala Met Cys Arg Ala Ser Leu Ala Met	
370 375 380	
TAT AAC ACC CAT GAA GAA GTG GAT CGT CTG GTG ACC GGC CTG CAA CGT	1200
Tyr Asn Thr His Glu Glu Val Asp Arg Leu Val Thr Gly Leu Gln Arg	
385 390 395 400	

31

1221

ATT CAC CGT TTG CTG GGA TAA
Ile His Arg Leu Leu Gly
405

(2) INFORMATION ZU SEQ ID NO: 8:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 406 Aminosäuren

(B) ART: Aminosäure

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 8:

Met	Ile	Phe	Ser	Val	Asp	Lys	Val	Arg	Ala	Asp	Phe	Pro	Val	Leu	Ser	
1				5					10					15		
Arg	Glu	Val	Asn	Gly	Leu	Pro	Leu	Ala	Tyr	Leu	Asp	Ser	Ala	Ala	Ser	
			20					25					30			
Ala	Gln	Lys	Pro	Ser	Gln	Val	Ile	Asp	Ala	Glu	Ala	Glu	Phe	Tyr	Arg	
		35					40					45				
His	Gly	Tyr	Ala	Ala	Val	His	Arg	Gly	Ile	His	Thr	Leu	Ser	Ala	Gln	
	50					55				60						
Ala	Thr	Glu	Lys	Met	Glu	Asn	Val	Arg	Lys	Arg	Ala	Ser	Leu	Phe	Ile	
65					70				75						80	
Asn	Ala	Arg	Ser	Ala	Glu	Glu	Leu	Val	Phe	Val	Arg	Gly	Thr	Thr	Glu	
			85						90					95		
Gly	Ile	Asn	Leu	Val	Ala	Asn	Ser	Trp	Gly	Asn	Ser	Asn	Val	Arg	Ala	
			100					105					110			
Gly	Asp	Asn	Ile	Ile	Ile	Ser	Gln	Met	Glu	His	His	Ala	Asn	Ile	Val	
		115					120					125				
Pro	Trp	Gln	Met	Leu	Cys	Ala	Arg	Val	Gly	Ala	Glu	Leu	Arg	Val	Ile	
	130					135					140					
Pro	Leu	Asn	Pro	Asp	Gly	Thr	Leu	Gln	Leu	Glu	Thr	Leu	Pro	Thr	Leu	
145					150					155					160	
Phe	Asp	Glu	Lys	Thr	Arg	Leu	Leu	Ala	Ile	Thr	His	Val	Ser	Asn	Val	
			165						170					175		
Leu	Gly	Thr	Glu	Asn	Pro	Leu	Ala	Glu	Met	Ile	Thr	Leu	Ala	His	Gln	
			180					185					190			
His	Gly	Ala	Lys	Val	Leu	Val	Asp	Gly	Ala	Gln	Ala	Val	Met	His	His	
	195						200						205			

32

Pro Val Asp Val Gln Ala Leu Asp Cys Asp Phe Tyr Val Phe Ser Gly
210 215 220

His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Ile Leu Tyr Val Lys Glu
225 230 235 240

Ala Leu Leu Gln Glu Met Pro Pro Trp Glu Gly Gly Gly Ser Met Ile
245 250 255

Ala Thr Val Ser Leu Ser Glu Gly Thr Thr Trp Thr Lys Ala Pro Trp
260 265 270

Arg Phe Glu Ala Gly Thr Pro Asn Thr Gly Gly Ile Ile Gly Leu Gly
275 280 285

Ala Ala Leu Glu Tyr Val Ser Ala Leu Gly Leu Asn Asn Ile Ala Glu
290 295 300

Tyr Glu Gln Asn Leu Met His Tyr Ala Leu Ser Gln Leu Glu Ser Val
305 310 315 320

Pro Asp Leu Thr Leu Tyr Gly Pro Gln Asn Arg Leu Gly Val Ile Ala
325 330 335

Phe Asn Leu Gly Lys His His Ala Tyr Asp Val Gly Ser Phe Leu Asp
340 345 350

Asn Tyr Gly Ile Ala Val Arg Thr Gly His His Cys Ala Met Pro Leu
355 360 365

Met Ala Tyr Tyr Asn Val Pro Ala Met Cys Arg Ala Ser Leu Ala Met
370 375 380

Tyr Asn Thr His Glu Glu Val Asp Arg Leu Val Thr Gly Leu Gln Arg
385 390 395 400

Ile His Arg Leu Leu Gly
405

(2) INFORMATION ZU SEQ ID NO: 9:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 3720 Basenpaare

(B) ART: Nukleinsäure

(C) STRANGFORM: Einzel

(D) TOPOLOGIE: ringförmig

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

33

(vii) UNMITTELBARE HERKUNFT:

(B) CLON: pHS1 metK

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: CDS

(B) LAGE: 530..1684

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 9:

GACGTCTGTG	TGGAATTGTG	AGCGGATAAC	AATTTACACAC	AGGGCCCTCG	GACACCGAGG	60
AGAATGTCAA	GAGGCGAACA	CACAACGTCT	TGGAGCGCCA	GAGGAGGAAC	GAGCTAAAAC	120
GGAGCTTTTT	TGCCCTGCGT	GACCAGATCC	CGGAGTTGGA	AAACAATGAA	AAGGCCCCCA	180
AGGTAGTTAT	CCTTAAAAAA	GCCACAGCAT	ACATCCTGTC	CGTCCAAGCA	GAGGAGCAAA	240
AGCTCATTTT	TGAAGAGGAC	TTGTTGCGGA	AACGACGAGA	ACAGTTGAAA	CACAAACTTG	300
ACAGCTACG	GAATCTTGT	GCGTAAGGAA	AAGTAAGGAA	AACGATTCCT	TCTAACAGAA	360
ATGTCCTGAG	CAATCACCTA	TGAACTGTCT	ACTCGAGATA	GCATTTTAT	CCATAAGATT	420
AGCCGATCCT	AAGGTTTACA	ATTGTGAGCG	CTCACAATTA	TGATAGATTC	AATTGTGAGC	480
GGATAACAAT	TTCACACACG	CTAGCGGTAC	CAAAGAGGAG	AAATTAACT	ATG GCA	535
				Met Ala		
				1		
AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT GAC AAA						583
Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro Asp Lys						
5	10	15				
ATT GCT GAC CAA ATT TCT GAT GCC GTT TTA GAC GCG ATC CTC GAA CAG						631
Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu Glu Gln						
20	25	30				
GAT CCG AAA GCA CGC GTT GCT TGC GAA ACC TAC GTA AAA ACC GGC ATG						679
Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr Gly Met						
35	40	45	50			
GTT TTA GTT GGC GGC GAA ATC ACC ACC AGC GCC TGG GTA GAC ATC GAA						727
Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp Ile Glu						
55	60	65				
GAG ATC ACC CGT AAC ACC GTT CGC GAA ATT GGC TAT GTG CAT TCC GAC						775
Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His Ser Asp						
70	75	80				
ATG GGC TTT GAC GCT AAC TCC TGT GCG GTT CTG AGC GCT ATC GGC AAA						823
Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile Gly Lys						
85	90	95				

34

CAG TCT CCT GAC ATC AAC CAG GGC GTT GAC CGT GCC GAT CCG CTG GAA Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro Leu Glu 100 105 110	871
CAG GGC GCG GGT GAC CAG GGT CTG ATG TTT GGC TAC GCA ACT AAT GAA Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr Asn Glu 115 120 125 130	919
ACC GAC GTG CTG ATG CCA GCA CCT ATC ACC TAT GCA CAC CGT CTG GTA Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg Leu Val 135 140 145	967
CAG CGT CAG GCT GAA GTG CGT AAA AAC GGC ACT CTG CCG TGG CTG CGC Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp Leu Arg 150 155 160	1015
CCG GAC GCG AAA AGC CAG GTG ACT TTT CAG TAT GAC GAC GGC AAA ATC Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly Lys Ile 165 170 175	1063
GTT GGT ATC GAT GCT GTC GTG CTT TCC ACT CAG CAC TCT GAA GAG ATC Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu Glu Ile 180 185 190	1111
GAC CAG AAA TCG CTG CAA GAA GCG GTA ATG GAA GAG ATC ATC AAG CCA Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile Lys Pro 195 200 205 210	1159
ATT CTG CCC GCT GAA TGG CTG ACT TCT GCC ACC AAA TTC TTC ATC AAC Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe Ile Asn 215 220 225	1207
CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC GGT CTG Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys Gly Leu 230 235 240	1255
ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG CGT CAC Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala Arg His 245 250 255	1303
GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC CGT TCC Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp Arg Ser 260 265 270	1351
GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT GCT GGC Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala Ala Gly 275 280 285 290	1399
CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC GTG GCT Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly Val Ala 295 300 305	1447

35

GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA GTG CCT	1495
Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys Val Pro	
310 315 320	
TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG CGC CCA	1543
Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu Arg Pro	
325 330 335	
TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC AAA GAA	1591
Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr Lys Glu	
340 345 350	
ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG GAA AAA	1639
Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp Glu Lys	
355 360 365 370	
ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG TAATCGGTAC	1691
Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys	
375 380 385	
CGCTTGATAT CGAATTCCTG CAGCCCGGGG GATCCCATGG TACGCGTGCT AGAGGCATCA	1751
AATAAAACGA AAGGCTCAGT CGAAAGACTG GGCCTTTCGT TTTATCTGTT GTTTGTCCGT	1811
GAACGCTCTC CTGAGTAGGA CAAATCCGCC GCCCTAGACC TAGGGGATAT ATTCCGCTTC	1871
CTCGCTCACT GACTCGCTAC GCTCGGTTCGT TCGACTGCGG CGAGCGGAAA TGGCTTACGA	1931
ACGGGGCGGA GATTTCTTGG AAGATGCCAG GAAGATACTT AACAGGGAAG TGAGAGGGCC	1991
GCGGCAAAGC CGTTTTTCCA TAGGCTCCGC CCCCTGACA AGCATCACGA AATCTGACGC	2051
TCAAATCAGT GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGC	2111
GGCTCCCTCG TGCGCTCTCC TGTTCCTGCC TTTCCGTTTA CCGGTGTCAT TCCGCTGTTA	2171
GGCCGCGTT TGTCTCATTC CACGCCTGAC ACTCAGTTCC GGGTAGGCAG TTCGCTCCAA	2231
GCTGGACTGT ATGCACGAAC CCCCCGTTCA GTCCGACCGC TGCGCCTTAT CCGGTAAC TA	2291
TCGTCTTGAG TCCAACCCGG AAAGACATGC AAAAGCACCA CTGGCAGCAG CCACTGGTAA	2351
TTGATTTAGA GGAGTTAGTC TTGAAGTCAT GCGCCGGTTA AGGCTAAACT GAAAGGACAA	2411
GTTTTGGTGA CTGCGCTCCT CCAAGCCAGT TACCTCGGTT CAAAGAGTTG GTAGCTCAGA	2471
GAACCTTCGA AAAACCGCCC TGCAAGGCGG TTTTTTCGTT TTCAGAGCAA GAGATTACGC	2531
GCAGACCAAA ACGATCTCAA GAAGATCATC TTATTAATCA GATAAAATAT TTCTAGATTT	2591
CAGTGCAATT TATCTCTTCA AATGTAGCAC CTGAAGTCAG CCCCATACGA TATAAGTTGT	2651
TACTAGTGCT TGGATTCTCA CCAATAAAAA ACGCCCGGCG GCAACCGAGC GTTCTGAACA	2711
AATCCAGATG GAGTTCTGAG GTCATTACTG GATCTATCAA CAGGAGTCCA AGCGAGCTCT	2771

36

CGAACCCAG AGTCCCGCTC AGAAGAACTC GTCAAGAAGG CGATAGAAGG CGATGCGCTG 2831
 CGAATCGGGA GCGGCGATAC CGTAAAGCAC GAGGAAGCGG TCAGCCCATT CGCCGCCAAG 2891
 CTCTTCAGCA ATATCACGGG TAGCCAACGC TATGTCCTGA TAGCGGTCCG CCACACCCAG 2951
 CCGGCCACAG TCGATGAATC CAGAAAAGCG GCCATTTTCC ACCATGATAT TCGGCAAGCA 3011
 GGCATCGCCA TGGGTCACGA CGAGATCCTC GCCGTCGGGC ATGCGCGCCT TGAGCCTGGC 3071
 GAACAGTTCG GCTGGCGCGA GCCCCTGATG CTCTTCGTCC AGATCATCCT GATCGACAAG 3131
 ACCGGCTTCC ATCCGAGTAC GTGCTCGCTC GATGCGATGT TTCGCTTGGT GGTCGAATGG 3191
 GCAGGTAGCC GGATCAAGCG TATGCAGCCG CCGCATTGCA TCAGCCATGA TGGATACTTT 3251
 CTCGGCAGGA GCAAGGTGAG ATGACAGGAG ATCCTGCCCC GGCCTTCGC CCAATAGCAG 3311
 TCAGTCCCTT CCCGCTTCAG TGACAACGTC GAGCACAGCT GCGCAAGGAA CGCCCGTCGT 3371
 GGCCAGCCAC GATAGCCGCG CTGCCTCGTC CTGCAGTTCA TTCAGGGCAC CGGACAGGTC 3431
 GGTCTTGACA AAAAGAACCG GGCGCCCCTG CGCTGACAGC CGGAACACGG CGGCATCAGA 3491
 GCAGCCGATT GTCTGTTGTG CCCAGTCATA GCCGAATAGC CTCTCCACCC AAGCGGCCGG 3551
 AGAACCTGCG TGCAATCCAT CTTGTTCAAT CATGCGAAAC GATCCTCATC CTGTCTCTTG 3611
 ATCAGATCTT GATCCCCTGC GCCATCAGAT CCTTGGCGGC AAGAAAGCCA TCCAGTTTAC 3671
 TTTGCAGGGC TTCCAACCT TACCAGAGGG CGCCCCAGCT GGCAATTCC 3720

(2) INFORMATION ZU SEQ ID NO: 10:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 384 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 10:

Met	Ala	Lys	His	Leu	Phe	Thr	Ser	Glu	Ser	Val	Ser	Glu	Gly	His	Pro
1				5					10					15	
Asp	Lys	Ile	Ala	Asp	Gln	Ile	Ser	Asp	Ala	Val	Leu	Asp	Ala	Ile	Leu
			20					25					30		
Glu	Gln	Asp	Pro	Lys	Ala	Arg	Val	Ala	Cys	Glu	Thr	Tyr	Val	Lys	Thr
		35					40					45			
Gly	Met	Val	Leu	Val	Gly	Gly	Glu	Ile	Thr	Thr	Ser	Ala	Trp	Val	Asp
	50					55					60				

37

Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
65 70 75 80

Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
85 90 95

Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
100 105 110

Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
115 120 125

Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
130 135 140

Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
145 150 155 160

Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
165 170 175

Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
180 185 190

Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
195 200 205

Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
210 215 220

Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
225 230 235 240

Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
245 250 255

Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
260 265 270

Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
275 280 285

Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
290 295 300

Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
305 310 315 320

Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
325 330 335

Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
340 345 350

38

Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
355 360 365

Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
370 375 380

(2) INFORMATION ZU SEQ ID NO: 11:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 3794 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: ringförmig

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

(vii) UNMITTELBARE HERKUNFT:

(B) CLON: pHS1 bioS1

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: CDS
- (B) LAGE: 601..1806

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 11:

GACGTCTGTG TGGAATTGTG AGCGGATAAC AATTTACAC AGGGCCCTCG GACACCGAGG 60
AGAATGTCAA GAGGCGAACA CACAACGTCT TGGAGCGCCA GAGGAGGAAC GAGCTAAAAC 120
GGAGCTTTTT TGCCCTGCGT GACCAGATCC CGGAGTTGGA AAACAATGAA AAGGCCCCCA 180
AGGTAGTTAT CCTTAAAAA GCCACAGCAT ACATCCTGTC CGTCCAAGCA GAGGAGCAAA 240
AGCTCATTTT TGAAGAGGAC TTGTGCGGA AACGACGAGA ACAGTTGAAA CACAAACTTG 300
AACAGCTACG GAACTCTTGT GCGTAAGGAA AAGTAAGGAA AACGATTCCT TCTAACAGAA 360
ATGTCCTGAG CAATCACCTA TGAAGTGTG ACTCGAGATA GCATTTTTAT CCATAAGATT 420
AGCCGATCCT AAGGTTTACA ATTGTGAGCG CTCACAATTA TGATAGATTC AATTGTGAGC 480
GGATAACAAT TTCACACACG CTAGCGGTAC CGGGCCCCC CTCGAGGTCG ACGGTATCGA 540
TAAGCTTGAT ATCGAATTCC TGCAGCCCGG GGGATCCCAT GGTACGCGTC GAGGAGTACC 600
ATG AAC GTT TTT AAT CCC GCG CAG TTT CGC GCC CAG TTT CCC GCA CTA 648
Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu

1

5

10

15

39

CAG	GAT	GCG	GGC	GTC	TAT	CTC	GAC	AGC	GCC	GCG	ACC	GCG	CTT	AAA	CCT	696
Gln	Asp	Ala	Gly	Val	Tyr	Leu	Asp	Ser	Ala	Ala	Thr	Ala	Leu	Lys	Pro	
			20					25					30			
GAA	GCC	GTG	GTT	GAA	GCC	ACC	CAA	CAG	TTT	TAC	AGT	CTG	AGC	GCC	GGA	744
Glu	Ala	Val	Val	Glu	Ala	Thr	Gln	Gln	Phe	Tyr	Ser	Leu	Ser	Ala	Gly	
		35					40					45				
AAC	GTC	CAT	CGC	AGC	CAG	TTT	GCC	GAA	GCC	CAA	CGC	CTG	ACC	GCG	CGT	792
Asn	Val	His	Arg	Ser	Gln	Phe	Ala	Glu	Ala	Gln	Arg	Leu	Thr	Ala	Arg	
		50				55					60					
TAT	GAA	GCT	GCA	CGA	GAG	AAA	GTG	GCG	CAA	TTA	CTG	AAT	GCA	CCG	GAT	840
Tyr	Glu	Ala	Ala	Arg	Glu	Lys	Val	Ala	Gln	Leu	Leu	Asn	Ala	Pro	Asp	
		65				70				75					80	
CAT	AAA	ACT	ATC	GTC	TGG	ACG	CGC	GGC	ACC	ACT	GAA	TCC	ATC	AAC	ATG	888
Asp	Lys	Thr	Ile	Val	Trp	Thr	Arg	Gly	Thr	Thr	Glu	Ser	Ile	Asn	Met	
				85					90					95		
GTG	GCA	CAA	TGC	TAT	GCG	CGT	CCG	CGT	CTG	CAA	CCG	GGC	GAT	GAG	ATT	936
Val	Ala	Gln	Cys	Tyr	Ala	Arg	Pro	Arg	Leu	Gln	Pro	Gly	Asp	Glu	Ile	
			100					105					110			
ATT	GTC	AGC	GTG	GCA	GAA	CAC	CAC	GCC	AAC	CTC	GTC	CCC	TGG	CTG	ATG	984
Ile	Val	Ser	Val	Ala	Glu	His	His	Ala	Asn	Leu	Val	Pro	Trp	Leu	Met	
		115					120					125				
GTC	GCC	CAA	CAA	ACT	GGA	GCC	AAA	GTG	GTG	AAA	TTG	CCG	CTT	AAT	GCG	1032
Val	Ala	Gln	Gln	Thr	Gly	Ala	Lys	Val	Val	Lys	Leu	Pro	Leu	Asn	Ala	
		130					135				140					
CAG	CGA	CTG	CCG	GAT	GTC	GAT	TTG	TTG	CCA	GAA	CTG	ATT	ACT	CCC	CGT	1080
Gln	Arg	Leu	Pro	Asp	Val	Asp	Leu	Leu	Pro	Glu	Leu	Ile	Thr	Pro	Arg	
		145				150				155					160	
AGT	CGG	ATT	CTG	GCG	TTG	GGT	CAG	ATG	TCG	AAC	GTT	ACT	GGC	GGT	TGC	1128
Ser	Arg	Ile	Leu	Ala	Leu	Gly	Gln	Met	Ser	Asn	Val	Thr	Gly	Gly	Cys	
				165				170						175		
CCG	GAT	CTG	GCG	CGA	GCG	ATT	ACC	TTT	GCT	CAT	TCA	GCC	GGG	ATG	GTG	1176
Pro	Asp	Leu	Ala	Arg	Ala	Ile	Thr	Phe	Ala	His	Ser	Ala	Gly	Met	Val	
				180				185					190			
GTG	ATG	GTT	GAT	GGT	GCT	CAG	GGG	GCA	GTG	CAT	TTC	CCC	GCG	GAT	GTT	1224
Val	Met	Val	Asp	Gly	Ala	Gln	Gly	Ala	Val	His	Phe	Pro	Ala	Asp	Val	
		195					200					205				
CAG	CAA	CTG	GAT	ATT	GAT	TTC	TAT	GCT	TTT	TCA	GGT	CAC	AAA	CTG	TAT	1272
Gln	Gln	Leu	Asp	Ile	Asp	Phe	Tyr	Ala	Phe	Ser	Gly	His	Lys	Leu	Tyr	
		210				215					220					

40

GGC CCG ACA GGT ATC GGC GTG CTG TAT GGT AAA TCA GAA CTG CTG GAG	1320
Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu	
225 230 235 240	
GCG ATG TCG CCC TGG CTG GGC GGC GGC AAA ATG GTT CAC GAA GTG AGT	1368
Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser	
245 250 255	
TTT GAC GGC TTC ACG ACT CAA TCT GCG CCG TGG AAA CTG GAA GCT GGA	1416
Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly	
260 265 270	
ACG CCA AAT GTC GCT GGT GTC ATA GGA TTA AGC GCG GCG CTG GAA TGG	1464
Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp	
275 280 285	
CTG GCA GAT TAC GAT ATC AAC CAG GCC GAA AGC TGG AGC CGT AGC TTA	1512
Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu	
290 295 300	
GCA ACG CTG GCG GAA GAT GCG CTG GCG AAA CGT CCC GGC TTT CGT TCA	1560
Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser	
305 310 315 320	
TTC CGC TGC CAG GAT TCC AGC CTG CTG GCC TTT GAT TTT GCT GGC GTT	1608
Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val	
325 330 335	
CAT CAT AGC GAT ATG GTG ACG CTG CTG GCG GAG TAC GGT ATT GCC CTG	1656
His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu	
340 345 350	
CGG GCC GGG CAG CAT TGC GCT CAG CCG CTA CTG GCA GAA TTA GGC GTA	1704
Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val	
355 360 365	
ACC GGC ACA CTG CGC GCC TCT TTT GCG CCA TAT AAT ACA AAG AGT GAT	1752
Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp	
370 375 380	
GTG GAT GCG CTG GTG AAT GCC GTT GAC CGC GCG CTG GAA TTA TTG GTG	1800
Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val	
385 390 395 400	
GAT TAAACGCGTG CTAGAGGCAT CAAATAAAAC GAAAGGCTCA GTCGAAAGAC	1853
Asp	
TGGGCCTTTC GTTTTATCTG TTGTTTGTCG GTGAACGCTC TCCTGAGTAG GACAAATCCG	1913
CCGCCCTAGA CCTAGGGGAT ATATTCCGCT TCCTCGCTCA CTGACTCGCT ACGCTCGGTC	1973
GTTCGACTGC GGCGAGCGGA AATGGCTTAC GAACGGGGCG GAGATTCCT GGAAGATGCC	2033

41

AGGAAGATAC	TTAACAGGGA	AGTGAGAGGG	CCGCGGCAAA	GCCGTTTTTC	CATAGGCTCC	2093
GCCCCCTGA	CAAGCATCAC	GAAATCTGAC	GCTCAAATCA	GTGGTGGCGA	AACCCGACAG	2153
GACTATAAAG	ATACCAGGCG	TTTCCCCCTG	GCGGCTCCCT	CGTGCGCTCT	CCTGTTCTTG	2213
CCTTTCGGTT	TACCGGTGTC	ATTCCGCTGT	TATGGCCGCG	TTTGTCTCAT	TCCACGCCTG	2273
ACACTCAGTT	CCGGGTAGGC	AGTTCGCTCC	AAGCTGGACT	GTATGCACGA	ACCCCCGTT	2333
CAGTCCGACC	GCTGCGCCTT	ATCCGGTAAC	TATCGTCTTG	AGTCCAACCC	GGAAAGACAT	2393
GCAAAAGCAC	CACTGGCAGC	AGCCACTGGT	AATTGATTTA	GAGGAGTTAG	TCTTGAAGTC	2453
ATGCGCCGGT	TAAGGCTAAA	CTGAAAGGAC	AAGTTTTGGT	GACTGCGCTC	CTCCAAGCCA	2513
GTTACCTCGG	TTCAAAGAGT	TGGTAGCTCA	GAGAACCCTC	GAAAAACCGC	CCTGCAAGGC	2573
GGTTTTTTTCG	TTTTCAGAGC	AAGAGATTAC	GCGCAGACCA	AAACGATCTC	AAGAAGATCA	2633
TCTTATTAAT	CAGATAAAAT	ATTTCTAGAT	TTCAGTGCAA	TTTATCTCTT	CAAATGTAGC	2693
ACCTGAAGTC	AGCCCCATAC	GATATAAGTT	GTTACTAGTG	CTTGGATTCT	CACCAATAAA	2753
AAACGCCCGG	CGGCAACCGA	GCGTCTTGAA	CAAATCCAGA	TGGAGTTCTG	AGGTCATTAC	2813
TGGATCTATC	AACAGGAGTC	CAAGCGAGCT	CTCGAACCCC	AGAGTCCCGC	TCAGAAGAAC	2873
TCGTCAAGAA	GGCGATAGAA	GGCGATGCGC	TGCGAATCGG	GAGCGGCGAT	ACCGTAAAGC	2933
ACGAGGAAGC	GGTCAGCCCA	TTCGCCGCCA	AGCTCTTCAG	CAATATCACG	GGTAGCCAAC	2993
GCTATGTCCT	GATAGCGGTC	CGCCACACCC	AGCCGGCCAC	AGTCGATGAA	TCCAGAAAAG	3053
CGGCCATTTT	CCACCATGAT	ATTCGGCAAG	CAGGCATCGC	CATGGGTCAC	GACGAGATCC	3113
TCGCCGTCGG	GCATGCGCGC	CTTGAGCCTG	GCGAACAGTT	CGGCTGGCGC	GAGCCCCTGA	3173
TGCTCTTCGT	CCAGATCATC	CTGATCGACA	AGACCGGCTT	CCATCCGAGT	ACGTGCTCGC	3233
TCGATGCGAT	GTTTCGCTTG	GTGGTCGAAT	GGGCAGGTAG	CCGGATCAAG	CGTATGCAGC	3293
CGCCGCATTG	CATCAGCCAT	GATGGATACT	TTCTCGGCAG	GAGCAAGGTG	AGATGACAGG	3353
AGATCCTGCC	CCGGCACTTC	GCCCAATAGC	AGCCAGTCCC	TTCCCGCTTC	AGTGACAACG	3413
TCGAGCACAG	CTGCGCAAGG	AACGCCCGTC	GTGGCCAGCC	ACGATAGCCG	CGCTGCCTCG	3473
TCCTGCAGTT	CATTCAGGGC	ACCGGACAGG	TCGGTCTTGA	CAAAAAGAAC	CGGGCGCCCC	3533
TGCGCTGACA	GCCGGAACAC	GGCGGCATCA	GAGCAGCCGA	TTGTCTGTTG	TGCCCAGTCA	3593
TAGCCGAATA	GCCTCTCCAC	CCAAGCGGCC	GGAGAACCTG	CGTGCAATCC	ATCTTGTTCA	3653
ATCATGCGAA	ACGATCCTCA	TCCTGTCTCT	TGATCAGATC	TTGATCCCCT	GCGCCATCAG	3713
ATCCTTGCGG	GCAAGAAAGC	CATCCAGTTT	ACTTTGCAGG	GCTTCCCAAC	CTTACCAGAG	3773

42

GGCGCCCCAG CTGGCAATTC C

3794

(2) INFORMATION ZU SEQ ID NO: 12:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 401 Aminosäuren

(B) ART: Aminosäure

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 12:

Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
1 5 10 15

Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
20 25 30

Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
35 40 45

Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
50 55 60

Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
65 70 75 80

Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
85 90 95

Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
100 105 110

Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
115 120 125

Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
130 135 140

Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145 150 155 160

Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
165 170 175

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
180 185 190

Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
195 200 205

Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
210 215 220

43

Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240

Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255

Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275 280 285

Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300

Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320

Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335

His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350

Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365

Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
370 375 380

Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385 390 395 400

Asp

(2) INFORMATION ZU SEQ ID NO: 13:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 4975 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: ringförmig

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

(vii) UNMITTELBARE HERKUNFT:

(B) CLON: pHS1 metK bioS1

44

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: CDS

(B) LAGE: 1782..2987

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: CDS

(B) LAGE: 530..1684

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 13:

GACGTCTGTG TGGAATTGTG AGCGGATAAC AATTTACAC AGGGCCCTCG GACACCGAGG	60
AGAATGTCAA GAGGCGAACA CACAACGTCT TGGAGCGCCA GAGGAGGAAC GAGCTAAAAC	120
GGAGCTTTTT TGCCCTGCGT GACCAGATCC CGGAGTTGGA AAACAATGAA AAGGCCCCCA	180
AGGTAGTTAT CCTTAAAAAA GCCACAGCAT ACATCCTGTC CGTCCAAGCA GAGGAGCAAA	240
GCTCATTTC TGAAGAGGAC TTGTTGCGGA AACGACGAGA ACAGTTGAAA CACAAACTTG	300
AACAGCTACG GAACTCTTGT GCGTAAGGAA AAGTAAGGAA AACGATTCCT TCTAACAGAA	360
ATGTCCTGAG CAATCACCTA TGAAGTGTG ACTCGAGATA GCATTTTTAT CCATAAGATT	420
AGCCGATCCT AAGGTTTACA ATTGTGAGCG CTCACAATTA TGATAGATTC AATTGTGAGC	480
GGATAACAAT TTCACACACG CTAGCGGTAC CAAAGAGGAG AAATTAACAT ATG GCA	535
	Met Ala
	1
AAA CAC CTT TTT ACG TCC GAG TCC GTC TCT GAA GGG CAT CCT GAC AAA	583
Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro Asp Lys	
5 10 15	
ATT GCT GAC CAA ATT TCT GAT GCC GTT TTA GAC GCG ATC CTC GAA CAG	631
Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu Glu Gln	
20 25 30	
GAT CCG AAA GCA CGC GTT GCT TGC GAA ACC TAC GTA AAA ACC GGC ATG	679
Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr Gly Met	
35 40 45 50	
GTT TTA GTT GGC GGC GAA ATC ACC ACC AGC GCC TGG GTA GAC ATC GAA	727
Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp Ile Glu	
55 60 65	
GAG ATC ACC CGT AAC ACC GTT CGC GAA ATT GGC TAT GTG CAT TCC GAC	775
Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His Ser Asp	
70 75 80	
ATG GGC TTT GAC GCT AAC TCC TGT GCG GTT CTG AGC GCT ATC GGC AAA	823
Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile Gly Lys	
85 90 95	

45

CAG TCT CCT GAC ATC AAC CAG GGC GTT GAC CGT GCC GAT CCG CTG GAA Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro Leu Glu 100 105 110	871
CAG GGC GCG GGT GAC CAG GGT CTG ATG TTT GGC TAC GCA ACT AAT GAA Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr Asn Glu 115 120 125 130	919
ACC GAC GTG CTG ATG CCA GCA CCT ATC ACC TAT GCA CAC CGT CTG GTA Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg Leu Val 135 140 145	967
CAG CGT CAG GCT GAA GTG CGT AAA AAC GGC ACT CTG CCG TGG CTG CGC Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp Leu Arg 150 155 160	1015
CCG GAC GCG AAA AGC CAG GTG ACT TTT CAG TAT GAC GAC GGC AAA ATC Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly Lys Ile 165 170 175	1063
GTT GGT ATC GAT GCT GTC GTG CTT TCC ACT CAG CAC TCT GAA GAG ATC Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu Glu Ile 180 185 190	1111
GAC CAG AAA TCG CTG CAA GAA GCG GTA ATG GAA GAG ATC ATC AAG CCA Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile Lys Pro 195 200 205 210	1159
ATT CTG CCC GCT GAA TGG CTG ACT TCT GCC ACC AAA TTC TTC ATC AAC Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe Ile Asn 215 220 225	1207
CCG ACC GGT CGT TTC GTT ATC GGT GGC CCA ATG GGT GAC TGC GGT CTG Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys Gly Leu 230 235 240	1255
ACT GGT CGT AAA ATT ATC GTT GAT ACC TAC GGC GGC ATG GCG CGT CAC Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala Arg His 245 250 255	1303
GGT GGC GGT GCA TTC TCT GGT AAA GAT CCA TCA AAA GTG GAC CGT TCC Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp Arg Ser 260 265 270	1351
GCA GCC TAC GCA GCA CGT TAT GTC GCG AAA AAC ATC GTT GCT GCT GGC Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala Ala Gly 275 280 285 290	1399
CTG GCC GAT CGT TGT GAA ATT CAG GTT TCC TAC GCA ATC GGC GTG GCT Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly Val Ala 295 300 305	1447

46

GAA CCG ACC TCC ATC ATG GTA GAA ACT TTC GGT ACT GAG AAA GTG CCT	1495
Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys Val Pro	
310 315 320	
TCT GAA CAA CTG ACC CTG CTG GTA CGT GAG TTC TTC GAC CTG CGC CCA	1543
Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu Arg Pro	
325 330 335	
TAC GGT CTG ATT CAG ATG CTG GAT CTG CTG CAC CCG ATC TAC AAA GAA	1591
Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr Lys Glu	
340 345 350	
ACC GCA GCA TAC GGT CAC TTT GGT CGT GAA CAT TTC CCG TGG GAA AAA	1639
Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp Glu Lys	
355 360 365 370	
ACC GAC AAA GCG CAG CTG CTG CGC GAT GCT GCC GGT CTG AAG TAATCGGTAC	1691
Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys	
375 380 385	
CGGGCCCCC CTCGAGGTCG ACGGTATCGA TAAGCTTGAT ATCGAATTCC TGCAGCCCCG	1751
GGGATCCCAT GGTACGCGTC GAGGAGTACC ATG AAC GTT TTT AAT CCC GCG CAG	1805
Met Asn Val Phe Asn Pro Ala Gln	
1 5	
TTT CGC GCC CAG TTT CCC GCA CTA CAG GAT GCG GGC GTC TAT CTC GAC	1853
Phe Arg Ala Gln Phe Pro Ala Leu Gln Asp Ala Gly Val Tyr Leu Asp	
10 15 20	
AGC GCC GCG ACC GCG CTT AAA CCT GAA GCC GTG GTT GAA GCC ACC CAA	1901
Ser Ala Ala Thr Ala Leu Lys Pro Glu Ala Val Val Glu Ala Thr Gln	
25 30 35 40	
CAG TTT TAC AGT CTG AGC GCC GGA AAC GTC CAT CGC AGC CAG TTT GCC	1949
Gln Phe Tyr Ser Leu Ser Ala Gly Asn Val His Arg Ser Gln Phe Ala	
45 50 55	
GAA GCC CAA CGC CTG ACC GCG CGT TAT GAA GCT GCA CGA GAG AAA GTG	1997
Glu Ala Gln Arg Leu Thr Ala Arg Tyr Glu Ala Ala Arg Glu Lys Val	
60 65 70	
GCG CAA TTA CTG AAT GCA CCG GAT GAT AAA ACT ATC GTC TGG ACG CGC	2045
Ala Gln Leu Leu Asn Ala Pro Asp Asp Lys Thr Ile Val Trp Thr Arg	
75 80 85	
GGC ACC ACT GAA TCC ATC AAC ATG GTG GCA CAA TGC TAT GCG CGT CCG	2093
Gly Thr Thr Glu Ser Ile Asn Met Val Ala Gln Cys Tyr Ala Arg Pro	
90 95 100	
CGT CTG CAA CCG GGC GAT GAG ATT ATT GTC AGC GTG GCA GAA CAC CAC	2141
Arg Leu Gln Pro Gly Asp Glu Ile Ile Val Ser Val Ala Glu His His	
105 110 115 120	

47

GCC AAC CTC GTC CCC TGG CTG ATG GTC GCC CAA CAA ACT GGA GCC AAA	2189
Ala Asn Leu Val Pro Trp Leu Met Val Ala Gln Gln Thr Gly Ala Lys	
125 130 135	
GTG GTG AAA TTG CCG CTT AAT GCG CAG CGA CTG CCG GAT GTC GAT TTG	2237
Val Val Lys Leu Pro Leu Asn Ala Gln Arg Leu Pro Asp Val Asp Leu	
140 145 150	
TTG CCA GAA CTG ATT ACT CCC CGT AGT CGG ATT CTG GCG TTG GGT CAG	2285
Leu Pro Glu Leu Ile Thr Pro Arg Ser Arg Ile Leu Ala Leu Gly Gln	
155 160 165	
ATG TCG AAC GTT ACT GGC GGT TGC CCG GAT CTG GCG CGA GCG ATT ACC	2333
Met Ser Asn Val Thr Gly Gly Cys Pro Asp Leu Ala Arg Ala Ile Thr	
170 175 180	
TTT GCT CAT TCA GCC GGG ATG GTG GTG ATG GTT GAT GGT GCT CAG GGG	2381
Phe Ala His Ser Ala Gly Met Val Val Met Val Asp Gly Ala Gln Gly	
185 190 195 200	
GCA GTG CAT TTC CCC GCG GAT GTT CAG CAA CTG GAT ATT GAT TTC TAT	2429
Ala Val His Phe Pro Ala Asp Val Gln Gln Leu Asp Ile Asp Phe Tyr	
205 210 215	
GCT TTT TCA GGT CAC AAA CTG TAT GGC CCG ACA GGT ATC GGC GTG CTG	2477
Ala Phe Ser Gly His Lys Leu Tyr Gly Pro Thr Gly Ile Gly Val Leu	
220 225 230	
TAT GGT AAA TCA GAA CTG CTG GAG GCG ATG TCG CCC TGG CTG GGC GGC	2525
Tyr Gly Lys Ser Glu Leu Leu Glu Ala Met Ser Pro Trp Leu Gly Gly	
235 240 245	
GGC AAA ATG GTT CAC GAA GTG AGT TTT GAC GGC TTC ACG ACT CAA TCT	2573
Gly Lys Met Val His Glu Val Ser Phe Asp Gly Phe Thr Thr Gln Ser	
250 255 260	
GCG CCG TGG AAA CTG GAA GCT GGA ACG CCA AAT GTC GCT GGT GTC ATA	2621
Ala Pro Trp Lys Leu Glu Ala Gly Thr Pro Asn Val Ala Gly Val Ile	
265 270 275 280	
GGA TTA AGC GCG GCG CTG GAA TGG CTG GCA GAT TAC GAT ATC AAC CAG	2669
Gly Leu Ser Ala Ala Leu Glu Trp Leu Ala Asp Tyr Asp Ile Asn Gln	
285 290 295	
GCC GAA AGC TGG AGC CGT AGC TTA GCA ACG CTG GCG GAA GAT GCG CTG	2717
Ala Glu Ser Trp Ser Arg Ser Leu Ala Thr Leu Ala Glu Asp Ala Leu	
300 305 310	
GCG AAA CGT CCC GGC TTT CGT TCA TTC CGC TGC CAG GAT TCC AGC CTG	2765
Ala Lys Arg Pro Gly Phe Arg Ser Phe Arg Cys Gln Asp Ser Ser Leu	
315 320 325	

CTG GCC TTT GAT TTT GCT GGC GTT CAT CAT AGC GAT ATG GTG ACG CTG	2813
Leu Ala Phe Asp Phe Ala Gly Val His His Ser Asp Met Val Thr Leu	
330 335 340	
CTG GCG GAG TAC GGT ATT GCC CTG CGG GCC GGG CAG CAT TGC GCT CAG	2861
Leu Ala Glu Tyr Gly Ile Ala Leu Arg Ala Gly Gln His Cys Ala Gln	
345 350 355 360	
CCG CTA CTG GCA GAA TTA GGC GTA ACC GGC ACA CTG CGC GCC TCT TTT	2909
Pro Leu Leu Ala Glu Leu Gly Val Thr Gly Thr Leu Arg Ala Ser Phe	
365 370 375	
GCG CCA TAT AAT ACA AAG AGT GAT GTG GAT GCG CTG GTG AAT GCC GTT	2957
Ala Pro Tyr Asn Thr Lys Ser Asp Val Asp Ala Leu Val Asn Ala Val	
380 385 390	
GAC CGC GCG CTG GAA TTA TTG GTG GAT TAAACGCGTG CTAGAGGCAT	3004
Asp Arg Ala Leu Glu Leu Leu Val Asp	
395 400	
CAAATAAAAC GAAAGGCTCA GTCGAAAGAC TGGGCCTTTC GTTTTATCTG TTGTTTGTCTG	3064
GTGAACGCTC TCCTGAGTAG GACAAATCCG CCGCCCTAGA CCTAGGGGAT ATATTCCGCT	3124
TCCTCGCTCA CTGACTCGCT ACGCTCGGTC GTTCGACTGC GGCGAGCGGA AATGGCTTAC	3184
GAACGGGGCG GAGATTTTCT GGAAGATGCC AGGAAGATAC TTAACAGGGA AGTGAGAGGG	3244
CCGCGGCAAA GCCGTTTTTC CATAGGCTCC GCCCCCTGA CAAGCATCAC GAAATCTGAC	3304
GCTCAAAATCA GTGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG	3364
GCGGCTCCCT CGTGCGCTCT CCTGTTCTCTG CCTTTCGGTT TACCGGTGTC ATTCCGCTGT	3424
TATGGCCGCG TTTGTCTCAT TCCACGCCTG AACTCAGTT CCGGGTAGGC AGTTCGCTCC	3484
AAGCTGGACT GTATGCACGA ACCCCCCGTT CAGTCCGACC GCTGCGCCTT ATCCGGTAAC	3544
TATCGTCTTG AGTCCAACCC GGAAAGACAT GCAAAAGCAC CACTGGCAGC AGCCACTGGT	3604
AATTGATTTA GAGGAGTTAG TCTTGAAGTC ATGCGCCGGT TAAGGCTAAA CTGAAAGGAC	3664
AAGTTTTGGT GACTGCGCTC CTCCAAGCCA GTTACCTCGG TTCAAAGAGT TGGTAGCTCA	3724
GAGAACCTTC GAAAAACCGC CCTGCAAGGC GGTTTTTTTCG TTTTCAGAGC AAGAGATTAC	3784
GCGCAGACCA AAACGATCTC AAGAAGATCA TCTTATTAAT CAGATAAAAT ATTTCTAGAT	3844
TTCAGTGCAA TTTATCTCTT CAAATGTAGC ACCTGAAGTC AGCCCCATAC GATATAAGTT	3904
GTTACTAGTG CTTGGATTCT CACCAATAAA AAACGCCCCG CGGCAACCGA GCGTTCTGAA	3964
CAAATCCAGA TGGAGTTCTG AGGTCATTAC TGGATCTATC AACAGGAGTC CAAGCGAGCT	4024
CTCGAACCCC AGAGTCCCGC TCAGAAGAAC TCGTCAAGAA GGCGATAGAA GGCGATGCGC	4084

49

TGCGAATCGG GAGCGGCGAT ACCGTAAAGC ACGAGGAAGC GGTCAGCCCA TTCGCCGCCA 4144
 AGCTCTTCAG CAATATCACG GGTAGCCAAC GCTATGTCCT GATAGCGGTC CGCCACACCC 4204
 AGCCGGCCAC AGTCGATGAA TCCAGAAAAG CGGCCATTTT CCACCATGAT ATTCGGCAAG 4264
 CAGGCATCGC CATGGGTCAC GACGAGATCC TCGCCGTCGG GCATGCGCGC CTTGAGCCTG 4324
 GCGAACAGTT CGGCTGGCGC GAGCCCCTGA TGCTCTTCGT CCAGATCATC CTGATCGACA 4384
 AGACCGGCTT CCATCCGAGT ACGTGCTCGC TCGATGCGAT GTTTCGCTTG GTGGTCGAAT 4444
 GGGCAGGTAG CCGGATCAAG CGTATGCAGC CGCCGCATTG CATCAGCCAT GATGGATACT 4504
 TTCTCGGCAG GAGCAAGGTG AGATGACAGG AGATCCTGCC CCGGCACTTC GCCCAATAGC 4564
 AGCCAGTCCC TTCCCCTTC AGTGACAACG TCGAGCACAG CTGCGCAAGG AACGCCCCTC 4624
 GTGGCCAGCC ACGATAGCCG CGCTGCCTCG TCCTGCAGTT CATTCAGGGC ACCGGACAGG 4684
 TCGGTCTTGA CAAAAGAAC CGGGCGCCCC TCGCTGACA GCCGGAACAC GGCGGCATCA 4744
 GAGCAGCCGA TTGTCTGTTG TGCCCAGTCA TAGCCGAATA GCCTCTCCAC CCAAGCGGCC 4804
 GGAGAACCTG CGTGCAATCC ATCTTGTTCA ATCATGCGAA ACGATCCTCA TCCTGTCTCT 4864
 TGATCAGATC TTGATCCCCT GCGCCATCAG ATCCTTGCGG GCAAGAAAGC CATCCAGTTT 4924
 ACTTTGCAGG GCTTCCCAAC CTTACCAGAG GGCGCCCCAG CTGGCAATTC C 4975

(2) INFORMATION ZU SEQ ID NO: 14:

(i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 384 Aminosäuren

(B) ART: Aminosäure

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 14:

Met Ala Lys His Leu Phe Thr Ser Glu Ser Val Ser Glu Gly His Pro
 1 5 10 15
 Asp Lys Ile Ala Asp Gln Ile Ser Asp Ala Val Leu Asp Ala Ile Leu
 20 25 30
 Glu Gln Asp Pro Lys Ala Arg Val Ala Cys Glu Thr Tyr Val Lys Thr
 35 40 45
 Gly Met Val Leu Val Gly Gly Glu Ile Thr Thr Ser Ala Trp Val Asp
 50 55 60
 Ile Glu Glu Ile Thr Arg Asn Thr Val Arg Glu Ile Gly Tyr Val His
 65 70 75 80

50

Ser Asp Met Gly Phe Asp Ala Asn Ser Cys Ala Val Leu Ser Ala Ile
85 90 95

Gly Lys Gln Ser Pro Asp Ile Asn Gln Gly Val Asp Arg Ala Asp Pro
100 105 110

Leu Glu Gln Gly Ala Gly Asp Gln Gly Leu Met Phe Gly Tyr Ala Thr
115 120 125

Asn Glu Thr Asp Val Leu Met Pro Ala Pro Ile Thr Tyr Ala His Arg
130 135 140

Leu Val Gln Arg Gln Ala Glu Val Arg Lys Asn Gly Thr Leu Pro Trp
145 150 155 160

Leu Arg Pro Asp Ala Lys Ser Gln Val Thr Phe Gln Tyr Asp Asp Gly
165 170 175

Lys Ile Val Gly Ile Asp Ala Val Val Leu Ser Thr Gln His Ser Glu
180 185 190

Glu Ile Asp Gln Lys Ser Leu Gln Glu Ala Val Met Glu Glu Ile Ile
195 200 205

Lys Pro Ile Leu Pro Ala Glu Trp Leu Thr Ser Ala Thr Lys Phe Phe
210 215 220

Ile Asn Pro Thr Gly Arg Phe Val Ile Gly Gly Pro Met Gly Asp Cys
225 230 235 240

Gly Leu Thr Gly Arg Lys Ile Ile Val Asp Thr Tyr Gly Gly Met Ala
245 250 255

Arg His Gly Gly Gly Ala Phe Ser Gly Lys Asp Pro Ser Lys Val Asp
260 265 270

Arg Ser Ala Ala Tyr Ala Ala Arg Tyr Val Ala Lys Asn Ile Val Ala
275 280 285

Ala Gly Leu Ala Asp Arg Cys Glu Ile Gln Val Ser Tyr Ala Ile Gly
290 295 300

Val Ala Glu Pro Thr Ser Ile Met Val Glu Thr Phe Gly Thr Glu Lys
305 310 315 320

Val Pro Ser Glu Gln Leu Thr Leu Leu Val Arg Glu Phe Phe Asp Leu
325 330 335

Arg Pro Tyr Gly Leu Ile Gln Met Leu Asp Leu Leu His Pro Ile Tyr
340 345 350

Lys Glu Thr Ala Ala Tyr Gly His Phe Gly Arg Glu His Phe Pro Trp
355 360 365

51

Glu Lys Thr Asp Lys Ala Gln Leu Leu Arg Asp Ala Ala Gly Leu Lys
370 375 380

(2) INFORMATION ZU SEQ ID NO: 15:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 401 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 15:

Met Asn Val Phe Asn Pro Ala Gln Phe Arg Ala Gln Phe Pro Ala Leu
1 5 10 15

Gln Asp Ala Gly Val Tyr Leu Asp Ser Ala Ala Thr Ala Leu Lys Pro
20 25 30

Glu Ala Val Val Glu Ala Thr Gln Gln Phe Tyr Ser Leu Ser Ala Gly
35 40 45

Asn Val His Arg Ser Gln Phe Ala Glu Ala Gln Arg Leu Thr Ala Arg
50 55 60

Tyr Glu Ala Ala Arg Glu Lys Val Ala Gln Leu Leu Asn Ala Pro Asp
65 70 75 80

Asp Lys Thr Ile Val Trp Thr Arg Gly Thr Thr Glu Ser Ile Asn Met
85 90 95

Val Ala Gln Cys Tyr Ala Arg Pro Arg Leu Gln Pro Gly Asp Glu Ile
100 105 110

Ile Val Ser Val Ala Glu His His Ala Asn Leu Val Pro Trp Leu Met
115 120 125

Val Ala Gln Gln Thr Gly Ala Lys Val Val Lys Leu Pro Leu Asn Ala
130 135 140

Gln Arg Leu Pro Asp Val Asp Leu Leu Pro Glu Leu Ile Thr Pro Arg
145 150 155 160

Ser Arg Ile Leu Ala Leu Gly Gln Met Ser Asn Val Thr Gly Gly Cys
165 170 175

Pro Asp Leu Ala Arg Ala Ile Thr Phe Ala His Ser Ala Gly Met Val
180 185 190

Val Met Val Asp Gly Ala Gln Gly Ala Val His Phe Pro Ala Asp Val
195 200 205

52

Gln Gln Leu Asp Ile Asp Phe Tyr Ala Phe Ser Gly His Lys Leu Tyr
210 215 220

Gly Pro Thr Gly Ile Gly Val Leu Tyr Gly Lys Ser Glu Leu Leu Glu
225 230 235 240

Ala Met Ser Pro Trp Leu Gly Gly Gly Lys Met Val His Glu Val Ser
245 250 255

Phe Asp Gly Phe Thr Thr Gln Ser Ala Pro Trp Lys Leu Glu Ala Gly
260 265 270

Thr Pro Asn Val Ala Gly Val Ile Gly Leu Ser Ala Ala Leu Glu Trp
275 280 285

Leu Ala Asp Tyr Asp Ile Asn Gln Ala Glu Ser Trp Ser Arg Ser Leu
290 295 300

Ala Thr Leu Ala Glu Asp Ala Leu Ala Lys Arg Pro Gly Phe Arg Ser
305 310 315 320

Phe Arg Cys Gln Asp Ser Ser Leu Leu Ala Phe Asp Phe Ala Gly Val
325 330 335

His His Ser Asp Met Val Thr Leu Leu Ala Glu Tyr Gly Ile Ala Leu
340 345 350

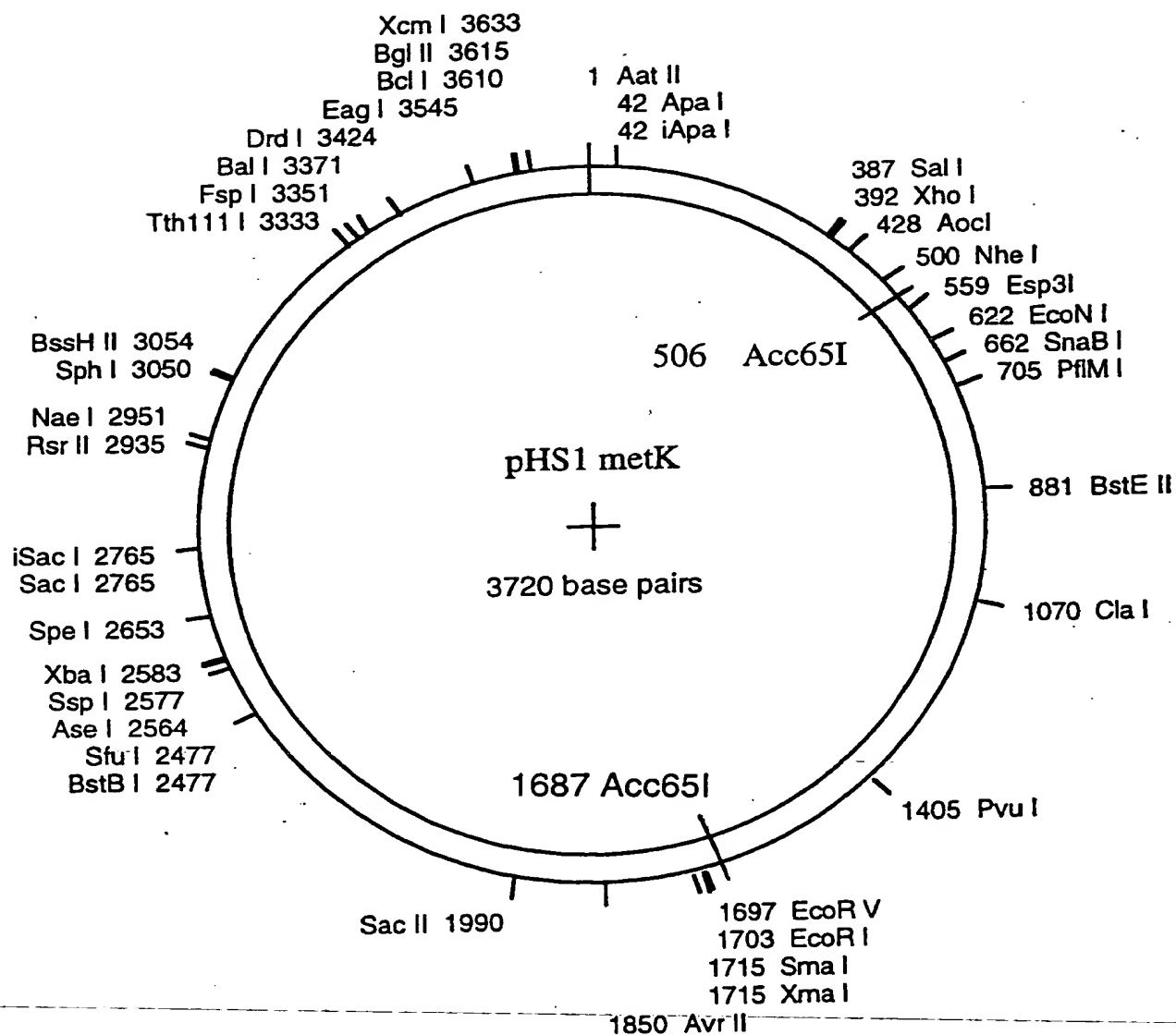
Arg Ala Gly Gln His Cys Ala Gln Pro Leu Leu Ala Glu Leu Gly Val
355 360 365

Thr Gly Thr Leu Arg Ala Ser Phe Ala Pro Tyr Asn Thr Lys Ser Asp
370 375 380

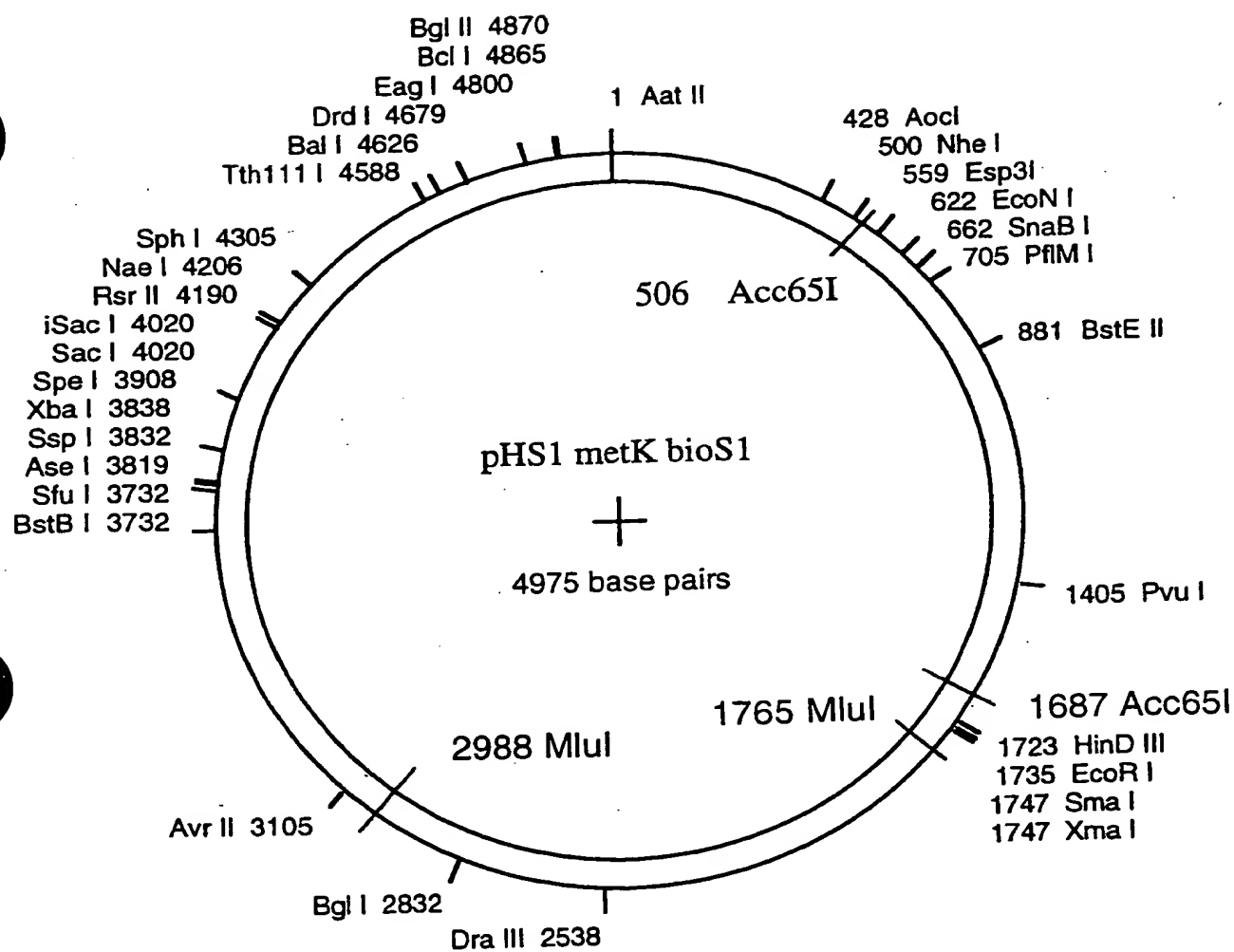
Val Asp Ala Leu Val Asn Ala Val Asp Arg Ala Leu Glu Leu Leu Val
385 390 395 400

sp

Figur 1



Figur 2



Verfahren zur Herstellung von Biotin

Zusammenfassung

5

Die Erfindung betrifft ein Genkonstrukt, enthaltend ein S-Adenosyl-Methionin-Synthasegen, mit der Sequenz SEQ ID No. 1 und ein Biotin-Biosynthesegen bioS1, bioS2 und/oder bioS3 mit den Sequenzen SEQ ID No.3, SEQ ID No.5 bzw. SEQ ID No.7 und gegebenenfalls
10 mindestens einer weiteren Biotinsynthesegensequenz ausgewählt aus der Gruppe bioA, bioB, bioF, bioC, bioD, bioH, bioP, bioW, bioX, bioY oder bioR. Die Erfindung betrifft weiterhin Organismen, die dieses Genkonstrukt enthalten und die Verwendung des Genkonstrukts zur Herstellung von Biotin, sowie ein Verfahren zur Her-
15 stellung von Biotin.

20

25

30

35

40

45

THIS PAGE BLANK (USPTO)